

IJCSIS Vol. 9 No. 5, May 2011
ISSN 1947-5500

International Journal of Computer Science & Information Security

© IJCSIS PUBLICATION 2011

Editorial Message from Managing Editor

The **Journal of Computer Science and Information Security (IJCSIS)** is a promising venue for publishing novel ideas, state-of-the-art research results and fundamental advances in all aspects of Computer Science and Engineering. **IJCSIS** is a peer reviewed international journal with a key objective to provide the academic and industrial community a medium for presenting high quality research. We are committed to timely publication of original research, surveying and tutorial contributions on the analysis and development of computing and information engineering. The journal is designed mainly to serve researchers and developers, dealing with information security and computing. Papers that can provide both theoretical analysis, along with carefully designed computational experiments, are particularly welcome.

IJCSIS editorial board consists of several internationally recognized experts and guest editors. Wide circulation is assured because libraries and individuals, worldwide, subscribe and reference to **IJCSIS**. The Journal has grown rapidly to its currently level of over 1,000 articles published and indexed; with distribution to librarians, universities, research centers, researchers in computing, and computer scientists.

Other field coverage includes: security infrastructures, network security: Internet security, content protection, cryptography, steganography and formal methods in information security; multimedia systems, software, information systems, intelligent systems, web services, data mining, wireless communication, networking and technologies, innovation technology and management. (See monthly Call for Papers)

IJCSIS is published using an open access publication model, meaning that all interested readers will be able to freely access the journal online without the need for a subscription. We wish to make **IJCSIS** one of the best journals in our area, i.e. Computer science in a wide sense, with growing impact factor.

On behalf of the Editorial Board and the **IJCSIS** members, we would like to express our gratitude to all authors and reviewers for their sustained support. The acceptance rate for this issue is 35%. I am confident that the readers of this journal will explore new avenues of research and academic excellence.

Available at <http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 9, No. 5, May 2011 Edition

ISSN 1947-5500 © IJCSIS, USA.

Abstracts Indexed by (among others):



IJCSIS EDITORIAL BOARD

Dr. M. Emre Celebi,

Assistant Professor, Department of Computer Science, Louisiana State University
in Shreveport, USA

Dr. Yong Li

School of Electronic and Information Engineering, Beijing Jiaotong University,
P. R. China

Prof. Hamid Reza Naji

Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran

Dr. Sanjay Jasola

Professor and Dean, School of Information and Communication Technology,
Gautam Buddha University

Dr Riktesh Srivastava

Assistant Professor, Information Systems, Skyline University College, University
City of Sharjah, Sharjah, PO 1797, UAE

Dr. Siddhivinayak Kulkarni

University of Ballarat, Ballarat, Victoria, Australia

Professor (Dr) Mokhtar Beldjehem

Sainte-Anne University, Halifax, NS, Canada

Dr. Alex Pappachen James, (Research Fellow)

Queensland Micro-nanotechnology center, Griffith University, Australia

Dr. T.C. Manjunath,

ATRIA Institute of Tech, India.

TABLE OF CONTENTS

1. Paper 21041109: Minimum Bit Error Rate Beamforming Combined with Space-Time Block Coding using Double Antenna Array Group (pp. 1-6)

Said Elnoubi, Electrical of Engineering, Alexandria University, Alexandria, Egypt
Waleed Abdallah, Tech. and App. Sc. Program, Al-Quds Open University, Jerusalem, Palestine
Mohamed M. M. Omar, Elect. & Comm. Eng., AAST, Abukir, Alexandria, Egypt

2. Paper 21041110: Analyzing and Comparing the Parsing Techniques of Asynchronous Message (pp. 7-12)

Mr. P. Krishna Sankar, Assistant Professor, Department of Computer science and Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi – 642 003
Ms. N. P. Shangaranarayane, Student, Department of Computer science and Engineering, Angel College of Engineering and Technology, Tirupur-641 665

3. Paper 21041113: Analysis on Differential Router Buffer Size towards Network Congestion (pp. 13-17)

Haniza N., Zulkiflee M., Abdul S. Shibgatullah, Shahrin S.
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

4. Paper 25041129: Mobility Assisted Solutions for Well-known Attacks in Mobile Wireless Sensor Network (pp. 18-22)

Abu Saleh Md. Tayeen¹, A.F.M. Sultanul Kabir¹
¹Department of Computer Science, American International University Bangladesh (AIUB), Dhaka, Bangladesh
Razib Hayat Khan, Department of Telematics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

5. Paper 26041131: Hybrid Multi-level Intrusion Detection System (pp. 23-29)

Sahar Selim, Mohamed Hashem and Taymoor M. Nazmy
Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

6. Paper 27041135: Skin Lesion Segmentation Algorithms using Edge Detectors (pp. 30-36)

J. H. Jaseema Yasmin¹, Associate Professor, Department of Computer Science and Engineering, National College of Engineering, Tirunelveli, India.
M. Mohamed Sathik², Associate Professor in Computer Science, Sadakathullah Appa College, Tirunelveli – India

7. Paper 27041137: Query Data With Fuzzy Information In Object-Oriented Databases An Approach The Semantic Neighborhood Of Hedge Algebras (pp. 37-42)

Doan Van Thang, Korea-VietNam Friendship Information Technology College, Department of Information systems, Faculty of Computer Science Da Nang City, Viet Nam
Doan Van Ban, Institute of Information Technology, Academy Science and Technology of Viet Nam, Ha Noi City, Viet Nam

8. Paper 29041145: A Low-Power CMOS Programmable CNN Cell and its Application to Stability of CNN with Opposite-Sign Templates (pp. 43-47)

S. El-Din, A. K. Abol Seoud, and A. El-Fahar

Electrical Engineering Department, University of Alexandria, Alexandria, Egypt.

M. El-Sayed Ragab, School of Electronics, Comm. and Computer Eng., E-JUST., Alexandria, Egypt.

9. Paper 30041148: A novel model for Synchronization and Positioning by using Neural Networks (pp. 48-53)

*Hossein Ghayoumi Zadeh *, Siamak Janianpour and Javad Haddadnia*

Department of Electrical Engineering, Sabzevar Tarbiat Moallem University, Sabzevar, Khorasan Razavi, Iran

10. Paper 30041149: Strategic Approach for Automatic Text Summarization (pp. 54-62)

Mr. Ramesh Vaishya, Sr. Lecturer, Department of Computer Science & Engg, Babu Banarsi Das National Institute of Technology & Management, Lucknow, India

Dr. Surya Prakash Tripathi, Associate Professor, Department of Computer Science & Engg, Institute of Engineering Technology, Lucknow, India

11. Paper 30041186: Golomb Ruler Sequences Optimization: A BBO Approach (pp. 63-71)

Shonak Bansal, Department of Electronics & Communications, Maharishi Markandeshwar University, Mullana, Haryana, India

Shakti Kumar, Computational Intelligence Laboratory, Institute of Science and Technology, Klawad, Haryana, India

Himanshu Sharma, Department of Electronics & Communications, Maharishi Markandeshwar University, Mullana, Haryana, India

Parvinder Bhalla, Computational Intelligence Laboratory, Institute of Science and Technology, Klawad, Haryana, India

12. Paper 21041108: Improving Enterprise Access Security Using RFID (pp. 72-77)

Dr. zakaria Saleh, Yarmouk University, Irbid, Jordan

Dr Izzat Alsmadi, Yarmouk University, Irbid, Jordan

Ahmed Mashhour Yarmouk University, Irbid, Jordan

13. Paper 21041112: Enhancement of stakeholders participations in Water fall Process Model (Step towards reducing the defects in software product) (pp. 78-80)

Mehar Ullah, Fasee Ullah, Muhammad Saeed Shehzad

Department of Computer Science, City University of Science & Information Technology (CUSIT),

Peshawar, Pakistan

14. Paper 21041115: Loopholes in Secure Socket layer and Sniffing (pp. 81-84)

Amit Mishra, Department of Computer Science & Engg., Faculty of Engineering & Technology, Jodhpur National University, Jodhpur, India

15. Paper 27041134: Secure Communication with Flipping Substitute Permutation Algorithm for Electronic Copy right Management System (pp. 85-94)

¹ C. Parthasarathy, ² G. Ramesh Kumar, ³ Dr. S. K. Srivatsa

¹ Sri Chandrashekhendra Saraswathi Viswa Mahavidyalaya University, Enathur, Kanchipuram – 631 561,

² Department of Computer Science & Applications, Adhiparasakthi College of Arts & Science, G. B. Nagar, Kalavai - 632 506, Vellore District. Tamil Nadu,

³ St. Joseph's College of Engg, Jeppiaar Nagar, Chennai-600 064

16. Paper 29041143: Context Based Word Sense Extraction in Text: Design Approach (pp. 95-99)

Ranjeetsingh S. Suryawanshi, Prof. D. M. Thakore, Kaustubh S. Raval

Bharati Vidyapeeth Deemed University, College of Engineering, Dhankawadi, Pune – 411043

17. Paper 30041153: An Overview on Radio Access Technology (RAT) Selection Algorithms for Heterogeneous Wireless Networks (pp. 100-105)

J. Preethi, Assistant Professor, Department of Computer Science and Engineering, Anna University of Technology, Coimbatore, India

Dr. S. Palaniswami, Professor, Department of Electrical and Electronics Engineering, Government College of Technology, India

18. Paper 30041156: Interactive Information System For Online Processing Geo-Technological Data (GTD) Sinking Wells (pp. 106-108)

Safarini Osama, IT Department, University of Tabuk, Tabuk, KSA

19. Paper 30041155: Extended RR-scheduling algorithm (pp. 109-113)

Prof. Sunita Chand, Prof. Teshu Chaudhary, Manoj Kumar

Krishna Engineering College, 95-Loni Road, Near Mohan Nagar, U.P-201007, India

20. Paper 30041158: Enhancement of Throughput for Single Hop WPAN's using UWB- OFDM Physical layer (pp. 114-118)

Ch. Subrahmanyam, Department of ECE, Scient Institute of Technology, Hyderabad, India

K. Chennakesava Reddy, Department of ECE, TKR College of Engg. & Tech., Hyderabad, India

Syed Abdul Sattar, Department of ECE, Royal Institute of Tech. & Science, Hyderabad, India

21. Paper 30041159: Enhancement of Throughput for Multi Hop WPAN's Using UWB - OFDM Physical layer (pp. 119-125)

Ch. Subrahmanyam, Department of ECE, Scient Institute of Technology, Hyderabad, India

K. Chennakesava Reddy, Department of ECE, TKR College of Engg. & Tech., Hyderabad, India

Syed Abdul Sattar, Department of ECE, Royal Institute of Tech. & Science, Hyderabad, India

22. Paper 30041162: Face Recognition Using Biogeography Based Optimization (pp. 126-131)

Er. Navdeep Kaur Johal, Er. Poonam Gupta, Er. Amandeep Kaur

Computer Science & Engineering, Rayat Institute of Engineering & Information Technology (RIEIT) Railmajra, Punjab, India

23. Paper 30041165: A Novel Steganographic Methodology For Secure Transmission Of Images (pp. 132-137)

B.V. Ramadevi, D. Lalitha Bhaskari, P.S. Avadhani

Department of Computer Science & Systems Engineering, AUCE(A), Andhra University, Visakhapatnam

24. Paper 30041188: A New Dynamic Data Allocation Algorithm for Distributed Database (pp. 138-141)

Fardin Esmaeeli Sangari, Sama Technical and vocational training college, Islamic Azad university, Urmia branch, Urmia, Iran
Seyed Mostafa Mansourfar, Sama Technical and vocational training college, Islamic Azad university, Sahand branch, Sahand, Iran

25. Paper 30041172: Establishing an Effective Combat Strategy for Prevalent Cyber- Attacks (pp. 142-148)

Vivian Ogochukwu Nwaocha, University of Nigeria, Nsukka, Computer Science Department
Inyama H.C., University of Nigeria, Nsukka, Computer Science Department

26. Paper 21041114: Accurate And Efficient Crawling The Deep Web: Surfacing Hidden Value (pp. 149-153)

Suneet Kumar, Associate Professor, DIT Dehradun
Anuj Kumar Yadav, Assistant Professor, DIT Dehradun
Rakesh Bharti, Assistant Professor, DIT Dehradun
Rani Choudhary, Sr. Lecturer, DBIT Ghaziabad

27. Paper 22111012: Mobile Phone Augmented Reality Business Card (pp. 154-164)

Edmund Ng Giap Weng, Centre of Excellence for Semantic Technology and Augmented Reality, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
Behrang Parhizkar, Faculty of Information & Communication Technology, LIMKOKWING University of Creative technology Cyberjaya, Selangor, Malaysia
Teo Tzong Ren, Centre of Excellence for Semantic Technology and Augmented Reality, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
Arash Habibi Lashkari, Faculty of Information & Communication Technology, LIMKOKWING University of Creative technology Cyberjaya, Selangor, Malaysia

28. Paper 25041127: Evercookies: Extremely persistent cookies (pp. 165-167)

Mohd. Shadab Siddiqui, Deepanker Verma
UPTU: CS&E, SRMCEM, Lucknow, India

29. Paper 27041136: Potential Research into Spatial Cancer Database by Using Data Clustering Techniques (pp. 168-173)

N. Naga Saranya, Research Scholar (C.S), Karpagam University, Coimbatore-641021, Tamilnadu, India.
Dr. M. Hemalatha, Head, Department of Software Systems, Karpagam University, Coimbatore-641021, Tamilnadu, India

30. Paper 23111016: Augmented Reality For Museum Artifacts Visualization (pp. 174-185)

Edmund Ng Giap Weng, Centre of Excellence for Semantic Technology and Augmented Reality, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
Behrang Parhizkar, Faculty of Information & Communication Technology, LIMKOKWING University of Creative technology Cyberjaya, Selangor, Malaysia
Lina Chai Hsiao Ping, Centre of Excellence for Semantic Technology and Augmented Reality, Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak
Arash Habibi Lashkari, Faculty of Information & Communication Technology, LIMKOKWING University of Creative technology Cyberjaya, Selangor, Malaysia

31. Paper 28041140: Cancelable Biometrics - A Survey (pp. 186-195)

Indira Chakravarthy, Associate Professor, Dept of CSE, Geethanjali College of Engg & Technology , Hyderabad

Dr. VVSSS. Balaram, Professor & Head, Dept of Information Technology, Sreenidhi Institute of Science & Technology , Hyderabad

Dr. B. Eswara Reddy, Associate Professor & Head, Dept of CSE, Jawaharlal Nehru Technological University, Anantapur

32. Paper 30041196: Reusable Code for CSOA-Services: Handling Data Coupling and Content Coupling (pp. 196-199)

Shakeel Ahmad, Sheikh Muhammad Saqib, Muhammad Ahmad Jan, Muhammad Zubair Asghar and Bashir Ahmad

Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

33. Paper 30041189: Computing the Efficiency of a DMU with Stochastic Inputs and Outputs Using Basic DEA Models (pp. 200-204)

M. Nabahat, Sama technical and vocational training school, Islamic Azad university, Urmia branch, Urmia, Iran

F. Esmaeeli sangari, Sama technical and vocational training school, Islamic Azad university, Urmia branch, Urmia, Iran

S. M. Mansourfar, Sama technical and vocational training school, Islamic Azad university, Sahand branch, Sahand, Iran

34. Paper 30041195: Concentration on Business Values for SOA-Services: A Strategy for Service's Business Values and Scope (pp. 205-208)

Bashir Ahmad, Sheikh Muhammad Saqib, Muhammad Zubair Asghar, Muhammad Ahmad Jan and Shakeel Ahmad

Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

35. Paper 31011180: Load-Balancing Geographic Routing Algorithm (ELBGR) For Wireless Sensor Networks (pp. 209-218)

Nazia Perwaiz, Department of Computer Engineering, NUST College of Electrical & Mechanical Engineering, National University of Sciences & Technology, Islamabad, Pakistan.

Dr. Muhammad Younus Javed, Department of Computer Engineering, NUST College of Electrical & Mechanical Engineering, National University of Sciences & Technology, Islamabad, Pakistan.

36. Paper 30041194: Custom Software under the Shade of Cloud Computing (pp. 219-223)

Sheikh Muhammad Saqib, Muhammad Ahmad Jan, Bashir Ahmad, Shakeel Ahmad and Muhammad Zubair Asghar

Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

37. Paper 27041118: Performance Analysis of The RAOA Protocol With Three Routing Protocols For Various Routing Metrics (pp. 224-230)

Lt. Dr. S. Santhosh Baboo, D G Vaishnav College, Arumbakkam, Chennai – 106.

V. J. Chakravarthy, Dravidian University

38. Paper 30041152: Multimedia Design Issues for Internet Telephony Protocols in Current High Performance Networks (pp. 231-239)

A. Jayachandran ,Asst. Professor, CSE dept, PSN College of Engineering and Technology, Tirunelveli, India

Dr. R. Dhanasekaran, Principal , Syed Ammal Engineering college, Ramnad, India
P. Rajan, Professor in MCA dept, PSN College of Engineering and Technology, Tirunelveli, India

39. Paper 30041174: Framework for Customized-SOA Projects (pp. 240-243)

Sheikh Muhammad Saqib, Muhammad Zubair Asghar, Shakeel Ahmad, Bashir Ahmad and Muhammad Ahmad Jan
Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

40. Paper 30041163: A New Hybrid Neural Model for Real-Time Prediction Applications (pp. 244-254)

Hazem M. El-Bakry[†] and Wael A. Awad^{††},
[†] Faculty of Computer Science & Information Systems, Mansoura University, Egypt
^{††} Math. & Comp. Dept., Faculty of Science, Port Said University, Port Said, Egypt

41. Paper 30041161: Energy Efficient Cluster Head Election using Fuzzy Logic in Wireless Sensor Networks (pp. 255-260)

Mostafa Basirnezhad, Department of Computer Engineering, Islamic Azad University of Mashhad, Mashhad, Iran
Dr. Masoud Niazi Torshiz, Department of Computer Engineering, Islamic Azad University of Mashhad, Mashhad, Iran

42. Paper 30041169: Effective Formal Procedure of Alternate Routings in MANET Improving Quality of Service (pp. 261-267)

Shakeel Ahmed and A. K. Ramani
School of Computer Science and Information Technology, Devi Ahilya University, Indore, India
Nazir Ahmad Zafar, Department of Computer Science, King Faisal University, Hofuf, Saudi Arabia

43. Paper 31121066: Speed Response and Performance Degradation of High Temperature Gamma Irradiated Silicon PIN Photodiodes (pp. 268-275)

Abd El-Naser A. Mohamed (1), Nabil A. Ayad (2), Ahmed Nabih Zaki Rashed (1) and Hazem M. El-Hageen1, (2)*
(1) Electronics and Electrical Communication Engineering Department, Faculty Electronic Engineering, Menouf, 32951, Egypt
(2) Atomic Energy Authority, P.O. Box 29, Naser City, Cairo, Egypt

44. Paper 30041176: A Multistage Detection and Elimination of Spurious Singular Points in Degraded Fingerprints (pp. 276-283)

Zia Saquib, Santosh Kumar Soni, Sweta Suhasaria
Center for Development of Advanced Computing, Mumbai, Maharashtra 400049, India
Dimple Parekh & Rekha Vig, NMIMS University, Mumbai, Maharashtra 400056, India

45. Paper 25041128: Practical Implementation Of Matlab Based Approach For Face Detection Using Feedforward Network (pp. 284-290)

Meenakshi Sharma, Sukhvinder Singh, Dr. N Suresh Rao
Sri Sai College Of Engg. & Tech., Pathankot, Jammu University

46. Paper 30041168: Reducing False Alerts Using Intelligent Hybrid Systems (pp. 291-297)

Sravan Kumar Jonnalagadda, Subha Sree Mallela

D.M.S.S.V.H. College of Engineering, Department of Information Technology, Machilipatnam, Andhra Pradesh, India

47. Paper 31041173: A Matlab Implementation of The Back-Propagation Approach for Reusability of Software Components (pp. 298-302)

*Meenakshi Sharma¹, Priyanka Kakkar², Dr. Parvinder Sandhu³
Sri Sai College Of Engg. & Tech., Pathankot^{1,2}, Rayat Kharar³*

48. Paper 30041157: Decision tree Induction Algorithm for Classification of Image Data (pp. 303-306)

*Kesari verma, Department of Computer Science, National Institute of Technology, Raipur, India
Ligendra Kumar Verma, Department of Computer Science, Raipur Institute of Technology, Raipur, India
Ajay Dewangan, Vivekananda College of Technology, Raipur, India*

Minimum Bit Error Rate Beamforming Combined with Space-Time Block Coding using Double Antenna Array Group

Said Elnoubi
Electrical of Engineering
Alexandria University
Alexandria, Egypt
saidelnoubi@hotmail.com

Waleed Abdallah
Tech. and App. Sc. Program
Al-Quds Open University,
Jerusalem, Palestine
wsalos@gou.edu

Mohamed M. M. Omar
Elect. & Comm. Eng.
AAST, Abukir
Alexandria, Egypt
mohammad_yosef@hotmail.com

Abstract— In this paper, we propose a Minimum Bit Error Rate (MBER) beamforming combined with Space-Time Block Coding (STBC) according to the number of antenna array. A class of adaptive beamforming algorithm has been proposed based on minimizing the BER cost function directly. Consequently, MBER beamforming is capable of providing significant performance gains in terms of a reduced BER. The beamforming weights of the combined system are optimized in such a way that the virtual channel coefficients corresponding to STBC-encoded data streams, seen at the receiver, are guaranteed to be uncorrelated. Therefore the promised achievable diversity order by conventional system with STBC can be obtained completely. Combined MBER beamforming with STBC single array performance measured by BER is compared under the condition of direction of arrival (DOA) and signal-to-noise ratio (SNR). The numerical simulation results of the proposed technique show that this minimum BER (MBER) approach utilizes the antenna array elements more intelligently and have a performance dependent of DOA and angular spread (AS).

Keywords—MBER beamforming; STBC; DOA; angular spread; adaptive antenna array

I. INTRODUCTION

The growing demand for wireless high-speed data transmission in applications such as wireless web browsing, file downloading, wireless multimedia transmission,..., etc., will increase requirements for downlink throughput and quality of service (QoS) significantly. But multipath fading is one of the major impairments limiting wireless communication systems in performance and capacity. Lots of new technologies such as smart antenna and transmit diversity have been proposed [1]. Those two technologies have the same features in the view of requiring the multiple antenna elements, but have the contradictive requirement for antenna element spacing.

Adaptive beamforming can separate signals transmitted in the same carrier frequency, provided that they are separated in the spatial domain. A beamformer combines the signal received by the different element of an antenna array to form a single output. This has been achieved by many criteria such as

maximizing SNR and minimizing the mean square error (MMSE) between the desired output and actual array output. This principle has its roots in the traditional beamforming employed in sonar and radar systems.

For a communication system, it is the achievable BER, not the MSE performance that really matters. Ideally, the system design should be based directly on minimizing the BER, rather than the MSE. For applications to single-user channel equalization and multi-user detection, it has been shown that the MMSE solution can in certain situations be distinctly inferior in comparison to the MBER solution, and several adaptive implementations of the MBER solution have been studied in the literature [3]. This contribution derives a novel adaptive beamforming technique based on directly minimizing the system's BER rather than the MSE. In [3], an adaptive implementation of the MBER beamforming technique is investigated.

STBC and beamforming techniques are two emerging technologies that can be employed at base station with multiple antennas to provide transmit diversity and beamforming gain to increase SNR of the downlink. In [1] and [2], the idea of the combination of two schemes to get the full diversity order as well as beamforming gain is proposed. There, the beamforming gain is achieved by maximizing received SNR at the receiver. It has shown real promise for increasing capacity and coverage and for mitigating multipath propagation of mobile radio communication systems.

In this paper, the MBER beamforming combined with STBC is proposed using single antenna array. This new technique is compared with the maximum SNR beamforming combined with STBC in array gain versus DOA center and BER versus DOA center and SNR performances. The simulation results show that the system's BER performance of the proposed algorithm is better than that investigated in [1], [2].

This paper is organized as follows. First, the combined beamforming with STBC single array is illustrated in Section II. Then the MBER beamforming algorithm is introduced in Section III. The combined MBER beamforming with STBC double array is presented in Section IV. In Section V, simulation results are conducted to evaluate the performance of

Identify applicable sponsor/s here. (sponsors)

the proposed scheme, the combined MBER beamforming with STBC single and double arrays, and compared with the performance of the combined maximum SNR with STBC single and double arrays followed by the conclusion in Section VI.

II. COMBINED BEAMFORMING WITH SPACE TIME BLOCK CODE SINGLE ARRAY

A. System Model

Fig. 1 shows the system employing STBC to combine with beamforming technique using single array [1-2]. The signal to be transmitted, $s(n)$, $1 \leq n \leq N$ is first coded using a STBC encoder, yielding two branch outputs as $s_1(n)$ and $s_2(n)$, where N is the number of transmitted bit sequences. They are then passed into two transmit beamformers w_1 and w_2 , respectively. At different time, they are simply added and transmitted as

$$x_1 = w_1^H \cdot s_1 + w_2^H \cdot s_2 \quad (1)$$

$$x_2 = w_1^H \cdot (-s_2^*) + w_2^H \cdot s_1^* \quad (2)$$

where w_i is the weight vector of the i^{th} beamformer and $(\cdot)^H$ is the Hermitian.

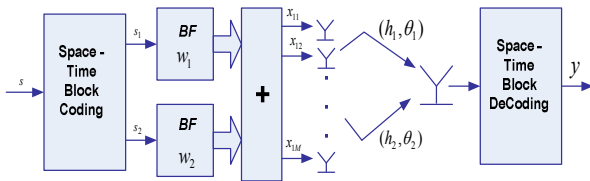


Figure 1. Combined beamforming with STBC using single array.

Suppose the physical channel consists of L spatially separated paths, whose fading coefficients and DOAs are denoted as (h_l, θ_l) for $l = 1 \dots L$. If the maximum time delay relative to the first arrived path is smaller than the symbol interval, a flat fading channel is observed and the instantaneous channel response can be expressed as

$$H = \sum_{l=1}^L h_l \cdot a(\theta_l) = \sum_{l=1}^L \alpha_l \exp(j\phi_l) \cdot a(\theta_l) \quad (3)$$

where α_l and ϕ_l are the fading amplitude and phase. For M -element uniform linear array (ULA) with spacing d , the downlink steering vector can be expressed as

$$a(\theta_l) = [1, e^{j2\pi \sin(\theta_l)d/\lambda}, \dots, e^{j2\pi(M-1)\sin(\theta_l)d/\lambda}]^T \quad (4)$$

So the received signal at the receiver is given by

$$r_1 = r(t) = w_1^H \cdot H \cdot s_1 + w_2^H \cdot H \cdot s_2 + \eta_1 \quad (5)$$

$$r_2 = r(t+T) = w_1^H \cdot H \cdot (-s_2^*) + w_2^H \cdot H \cdot (s_1^*) + \eta_2 \quad (6)$$

where T is the symbol duration, r_1 and r_2 are the received signals at time t and $t+T$, η_1 and η_2 are complex-valued white Gaussian noise having a zero mean and a variance of $2\sigma_\eta^2$.

B. Detection

In order to get maximal SNR, [1] tried to maximize (7) subject to (8) based on conventional STBC detection

$$E \left[|w_1^H \cdot H|^2 + |w_2^H \cdot H|^2 \right] \quad (7)$$

$$w_1^H \cdot w_1 + w_2^H \cdot w_2 = 1 \quad (8)$$

The downlink channel covariance matrix (DCCM) $E[H^H H]$ is well analyzed in [4] for TDD and FDD system.

For simplicity set $L=2$, then equations (5) and (6) can be rewritten as

$$r_1 = (w_1^H \cdot s_1 + w_2^H \cdot s_2) \cdot [h_1 \cdot a(\theta_1) + h_2 \cdot a(\theta_2)] + \eta_1 \quad (9)$$

$$r_2 = [(w_1^H \cdot (-s_2^*) + w_2^H \cdot (s_1^*)) \cdot [h_1 \cdot a(\theta_1) + h_2 \cdot a(\theta_2)] + \eta_2 \quad (10)$$

In [2], at receiver the Alamouti STBC (2Tx, 1Rx) [5] detection is used

$$\tilde{s}_1 = h_1^* \cdot r_1 + h_2 \cdot r_2^* \quad (11)$$

and the beamforming weight vectors w_1 and w_2 are set to be

$$w_1 = \frac{1}{\sqrt{2M}} \cdot a(\theta_1), \quad w_2 = \frac{1}{\sqrt{2M}} \cdot a(\theta_2) \quad (12)$$

which are maximizing the receiving SNR at the receiver.

The transmit beamforming weight are optimized by maximizing the cost function, but increasing the computing complexity [2].

III. MBER BEAMFORMING WITH STBC SOLUTION

It is assumed that the system supports L users, each user transmits signal on the same carrier frequency. The linear antenna array considered consists of M uniformly spaced elements and the signal received by the M -element antenna array are given by

$$x(n) = [a(\theta_1), a(\theta_2), \dots, a(\theta_L)] \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_L(n) \end{bmatrix} \quad (13)$$

where s_i is the signal to be transmitted for i^{th} user. $s_1(n)$ is assumed to be the desired user and the rest of the sources are the interfering users. To determine the MBER beamforming weight vector w , we start by forming its BER cost function [6]. The conditional probability density function (pdf) given by

$$P(y_s) = \frac{1}{N\sqrt{2\pi\sigma_\eta^2}} \sum_{n=1}^N \exp \left(-\frac{(y_s - \text{sgn}(s_1(n))y_R(n))^2}{2\sigma_\eta^2} \right) \quad (14)$$

is the best indicator of a beamformer's BER performance, where

$$y(n) = w^H x(n) \quad (15)$$

$$y_s(n) = \text{sgn}(s_1(n))y_R(n) \quad (16)$$

$\text{sgn}(\cdot)$ denotes the sign function, $y_R(n) = \text{Re}\{y(n)\}$ is the real part of the beamformer output $y(n)$ and $y_s(n)$ is an error indicator for the binary decision, i.e., if it is positive, then we

have a correct decision, else if it is negative, then an error occurred.

Hence, the error probability of the beamformer w , the BER cost function, is given by

$$P_E(w) = \frac{1}{N} \sum_{n=1}^N Q(g_n(w)) \quad (17)$$

where $Q(\cdot)$ is the Gaussian error function given by

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} \exp\left(-\frac{v^2}{2}\right) dv \quad (18)$$

and

$$g_n(w) = \frac{\text{sgn}(s_1(n))y_R(n)}{\sigma_\eta} \quad (19)$$

The MBER beamforming solution is then defined as

$$w_{MBER} = \arg \min_w P_E(w) \quad (20)$$

The gradient of $P_E(w)$ with respect to w can be shown to be

$$\nabla P_E(w) = \frac{\partial P_E(w)}{\partial w} = \frac{1}{2N\sqrt{2\pi\sigma_\eta^2}} \sum_{n=1}^N \exp\left(-\frac{(y_R(n))^2}{2\sigma_\eta^2}\right) \cdot \text{sgn}(s_1(n))(y_R(n)w - x(n)) \quad (21)$$

The following simplified conjugate gradient algorithm [3] provides an efficient means of finding a MBER solution.

In this paper, we will demonstrate from the simulation results that the system's BER performance can be improved by applying the MBER solutions instead of the beamforming weight vectors given by (11) combined with STBC.

The proposed MBER algorithm is summarized in Table I. We initialize the main algorithm parameters. The algorithm consists of one main loop. This loop is repeated until the norm of the gradient vector is sufficiently small.

1) Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. SUMMARY OF THE MBER ALGORITHM

Initialization
$w(0) = x(0)/\ x(0)\ $, $\mu = .8$, $\beta = .01$ (typically, β can be set to the machine accuracy). The adaptive gain μ and a termination scalar β are the two algorithmic parameters that have to be set appropriately to ensure a fast convergence rate and small steady-state BER. <ul style="list-style-type: none"> • Calculate variance of noise. • Calculate the gradient vector form (21). • Complexity of (21) is $O(M)$ for one bit [6]. • Initialize the search direction, $D = -\nabla P_E$, $i=1$; $\ \nabla P_E\$
<p style="text-align: center;"><i>while</i> ($\ \nabla P_E\ < \beta$)</p> <ul style="list-style-type: none"> • Update the beamformer weight $w(i+1) = w(i) + \mu D$ • Normalize the solution $w(i+1) = w(i+1)/\ w(i+1)\$ • Calculate the cost function BER and the gradient vector $\nabla P_E(w) = \frac{\partial P_E(w)}{\partial w} = \frac{1}{2N\sqrt{2\pi\sigma_\eta^2}} \sum_{n=1}^N \exp\left(-\frac{(y_R(n))^2}{2\sigma_\eta^2}\right) \cdot \text{sgn}(s_1(n))(y_R(n)w - x(n))$ <p style="text-align: center;">Complexity is $O(M)$ for one bit [6].</p>

$$P_E(w) = \frac{1}{N} \sum_{n=1}^N Q(g_n(w));$$

$$g_n(w) = \frac{\text{sgn}(s_1(n))y_R(n)}{\sigma_\eta};$$

- Calculate the search direction from

$$\phi_i = \frac{\|\nabla P_E(w(i+1))\|^2}{\|\nabla P_E(w(i))\|^2};$$

$$D(i+1) = \phi_i D(i) - \nabla P_E(w(i+1));$$

- Increment the iteration number $i = i + 1$
- end of while loop

Stop : $w(i)$ is the solution of the MBER weight vector.

To determine the MBER beamforming weight vector for another user, we can apply the algorithm stated in Table I for choosing $s_2(n)$ as desired user and the remainder of the sources are considered to be interfering sources.

As shown in [1], the equation denoted as array gain is given by

$$\mathcal{E} = \frac{\|w_2^H \cdot w_1\|^2}{\|w_2^H \cdot w_2\|^2} \quad (22)$$

Fig.2 shows the array gain depends on DOA (center) and angular spread (AS). At 10° AS case, as DOA (center) are 0° and 60° , \mathcal{E} are equal to 0.378 and 0.799 for the maximum SNR and are equal to 0.39 and 0.843 for the proposed algorithm, respectively. It changes widely enough to affect the performance for two algorithms.

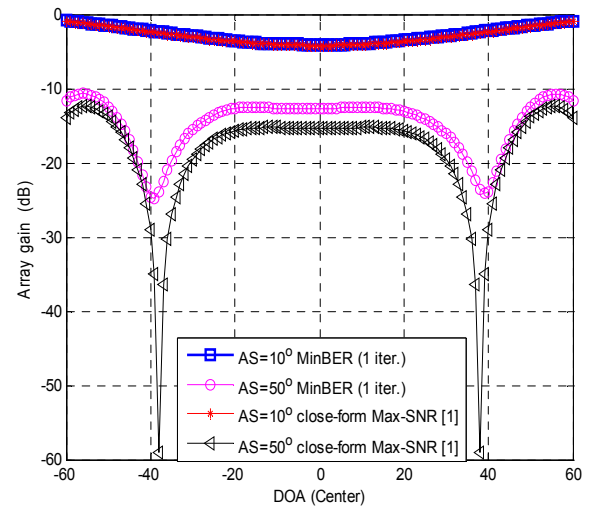


Figure 2. Array gain.

IV. COMBINED BEAMFORMING WITH SPACE TIME BLOCK CODE DOUBLE ARRAY

For array gain will strongly affect the system detection performance, we find another scheme to minimize the disadvantage.

Fig.3 shows the double array (Combined beamforming with space time block code double array) model. Unlike combined beamforming with space time block code single array model, after being put into the two beamformers, two data streams are sent by two dependent antenna arrays. The element number for one array is M . All parameters of equations shown in Fig.3 are same as those in section II.

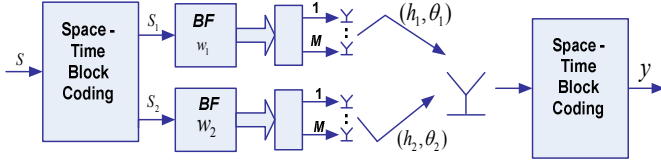


Figure 3. Combined beamforming with STBC using double array.

The received signals at the mobile terminal can be expressed as:

$$r_1 = w_1^H \cdot h_1 \cdot a(\theta_1) \cdot s_1 + w_2^H \cdot h_2 \cdot a(\theta_2) \cdot s_2 + \eta_1 \quad (23)$$

$$r_2 = w_1^H \cdot h_1 \cdot a(\theta_1) \cdot (-s_1^*) + w_2^H \cdot h_2 \cdot a(\theta_2) \cdot s_1^* + \eta_2 \quad (24)$$

And the detection for s_1 is

$$\tilde{s}_1 = h_1^* \cdot r_1 + h_2 \cdot r_2^* \quad (25)$$

V. SIMULATION RESULTS

In our numerical simulations, we consider the same example investigated in [1] to make comparisons. A 6-element uniform linear array (ULA) antenna is assumed in the base station with element spacing of $\lambda/2$, while the mobile terminal has single antenna. We simulate the BER supposing the desired user moves in a sector of 120° . The channel is assumed suffering from Rayleigh fading with various AS.

Fig.4 and Fig.5 illustrate the average BER performance of the combined beamforming with space time block coding (CB-STBC) single array using maximum SNR and MBER schemes versus DOA for two different cases, AS = 50° and 10° .

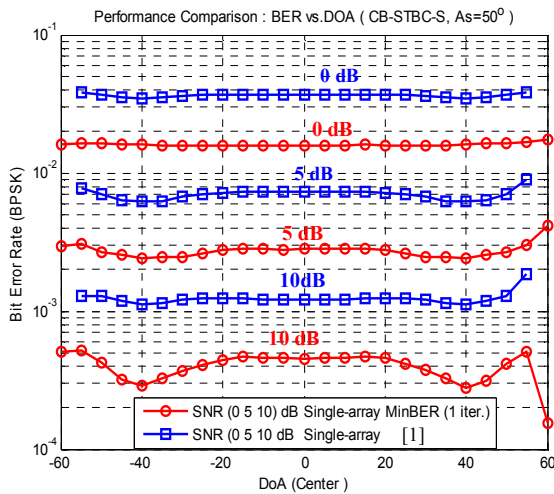


Figure 4. Performance comparison: BER vs. DOA (Combined beamforming with STBC using single array, AS= 50°).

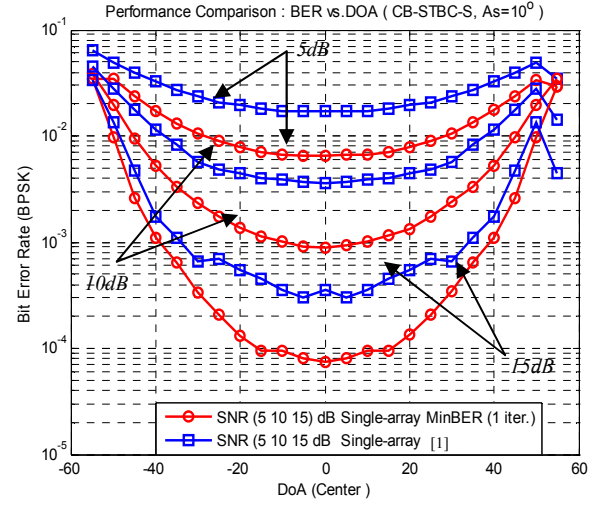


Figure 5. Performance comparison: BER vs. DOA (Combined beamforming with STBC using single array, AS= 10°).

It can be seen for large angular spread the BER performance does not affected by DOA but is seriously affected for small angular spread case, especially with bigger SNR.

Fig.6 and Fig.7 illustrate the average BER performance of the CB-STBC single array using maximum SNR and MBER schemes versus SNR. Also, the same two cases are considered in each Figure to represent the cases with small and large AS. For this example, the superior performance of the MBER scheme over the MSNR scheme becomes evident.

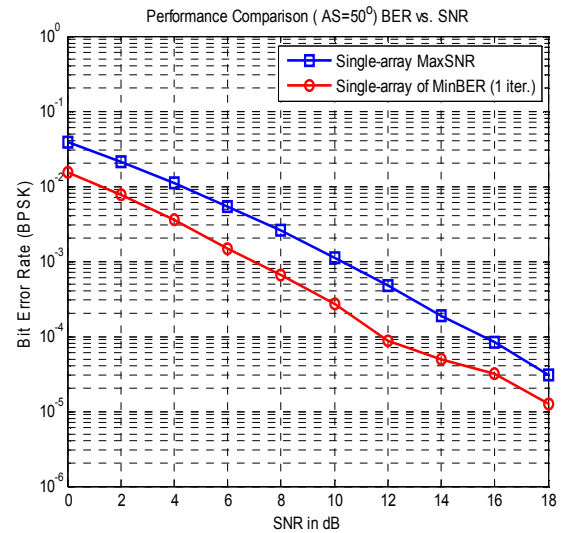


Figure 6: Performance Comparison: BER vs. SNR.

Combined beamforming with STBC using double array overcomes the disadvantages appeared on the single array model. Fig. 8 and 9 show us a stable performance which is not dependent on AS.

Fig.10 illustrates the average BER performance of the CB-STBC double array using maximum SNR and MBER schemes

versus SNR. Also, the same two cases are considered in each Figure to represent the cases with small and large AS.

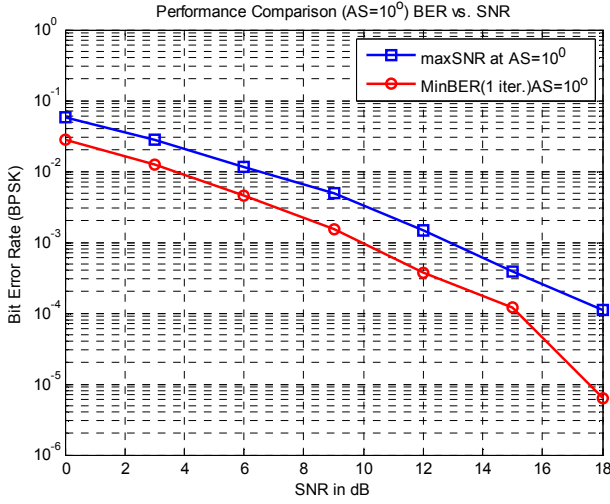


Figure 7: Performance Comparison: BER vs. SNR.

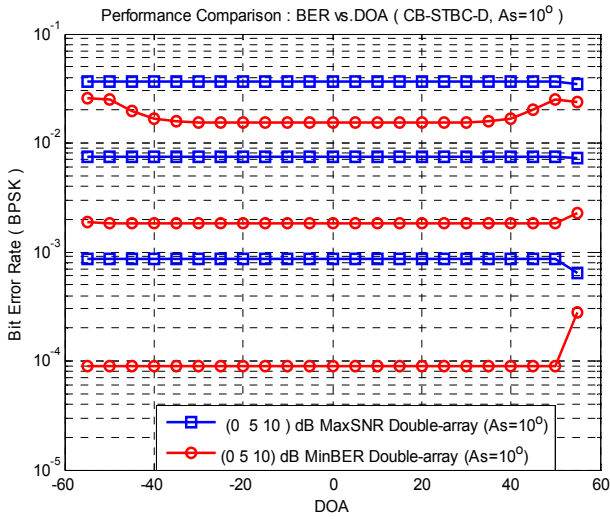


Figure 8. Performance comparison : BER vs. DOA (combined beamforming with STBC double array, $As=10^0$).

A. Computational Complexity

The proposed MBER maintains the linearity in complexity; however, its performance is better than the maximum SNR algorithm. Since addition is much easier than multiplication, we focus on multiplication complexities. Table I, illustrates the number of multiplication required to complete a single iteration, i.e., detecting one bit.

B. Convergent Rate

In this section, we run the algorithm of the MBER for 1000 samples and are limited to 1 and 11 iterations. The results are shown in Fig. 11, where we can see that the proposed algorithm converges very fast to the optimal solution (after one iteration only).

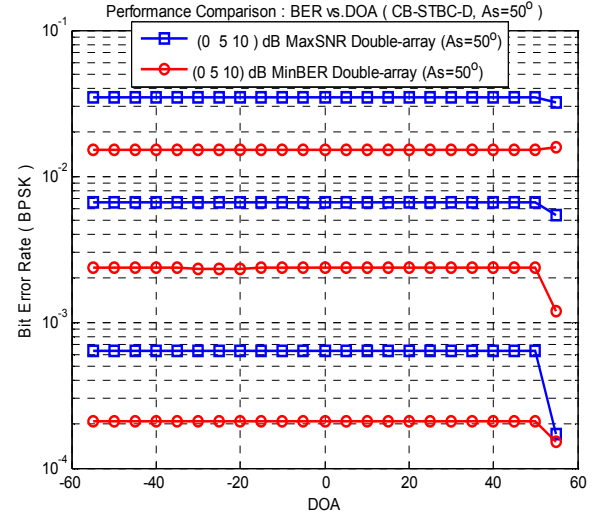


Figure 9. Performance comparison: BER vs. DOA (combined beamforming with STBC double array, $As=50^0$).

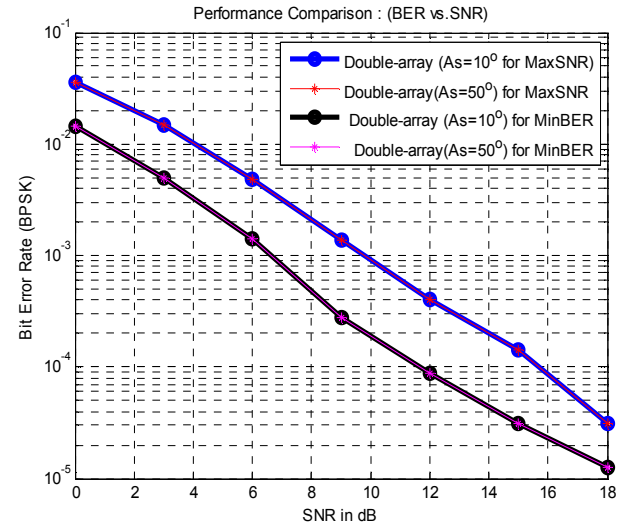


Figure 10. Performance Comparison: BER vs. SNR with $DOA(\text{center})=0^0$.

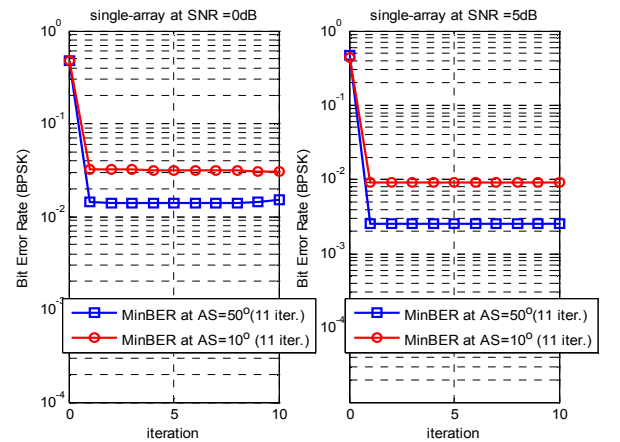


Figure 11. Convergence rate vs. iteration of the MBER algorithm.

Furthermore, we can observe in Fig. 12, a significant improvement over the maximum SNR algorithm by means of only one iteration.

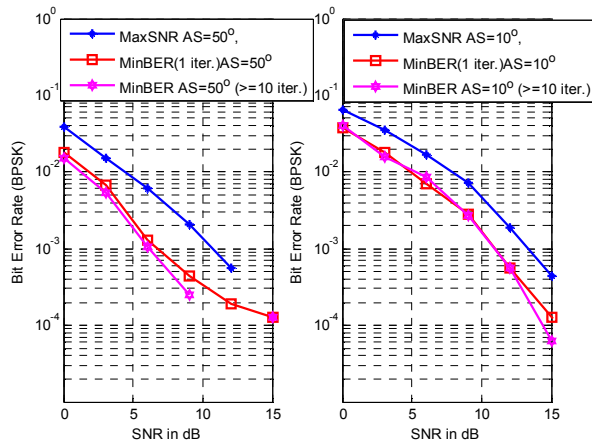


Fig.12. : Convergence of the MinBER algorithm.

VI. CONCLUSION

In this paper, a downlink transmit diversity scheme is proposed to achieve both full diversity gain and optimized beamforming gain. It is obtained by combining MBER beamforming technique with STBC for multiple beamforming antenna systems (single and double array). An adaptive MBER beamforming technique has been developed. It has been shown that the MBER beamformer exploits the system's resources more intelligently than the other standard beamformers and, consequently, can achieve a better performance in terms of a lower BER.

The combined beamforming with STBC using single array are shown to be dependent on the DOA and angular spread. However combined beamforming with STBC using double array is shown to have a stable performance independent of DOA and angular spread.

REFERENCES

- [1] Fan Zhu, Kyung Sik, Myoung Lim, "Combined beamforming with space-time block coding using double antenna array group", Electronics Letters, 2004, 40 (13):811-813.
- [2] Zhongding Lei, Chin F.P.S., Ying-Chang Liang, "Combined beamforming with space-time block coding for wireless downlink transmission," Vehicular Technology Conference, 2002. Proceedings .VTC 2002-Fall. 2002 IEEE 56th, Vol. 4, pp. 24-28 Sept. 2002.
- [3] Frank B.Gross, PhD, "Smart Antennas For Wireless Communications with matlab", 2005.
- [4] Said Elnoubi, Waleed Abdallah, Mohamed M. M. Omar, "Minimum Bit Error Rate Beamforming Combined with Space-time Block Coding", International Conference on Communications and Information Tech. ICCIT-2011, Aqaba-Jordan.
- [5] Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," IEEE Trans. Inf. Theory, vol. 44, no. 2, pp. 744-765, Mar. 1998.

- [6] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," IEEE Trans. Signal Processing, vol. 50, no. 10, pp. 2599-2613, Oct. 2002.
- [7] Y-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems", IEEE JSAC, Vol. 19, pp. 222-232, Feb. 2001.
- [8] M. Lin, M. Li, L. Yang and X. You, "Adaptive transmit beamforming with space-time coding for correlated MIMO fading channels," in Proc. IEEE ICC '07, June 2007.
- [9] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," IEEE JSAC, Vol. 16, No. 8, pp. 1451-1458, October 1998.
- [10] M. Lin, M. Li, L. Yang and X. You, "Adaptive transmit beamforming with space-time coding for correlated MIMO fading channels," in Proc. IEEE ICC '07, June 2007.
- [11] S. Chen, N. N. Ahmad, and L. Hanzo, "Adaptive Minimum Bit-Error Rate Beamforming", IEEE Transactions on Wireless Communications, VOL. 4, NO. 2 MARCH 2005.
- [12] T. A. Samir, S. Elnoubi, and A. Elnashar, "Class of minimum bit error rate algorithms," in Proc. 9th ICACT, Feb. 12-14, 2007, vol. 1, pp. 168-173.
- [13] S. Chen, A. K. Samangan, B. Mulgrew, and L. Hanzo, "Adaptive minimum- BER linear multiuser detection for DS-CDMA signals in multipath channels," IEEE Trans. Signal Process., vol. 49, no. 6, pp. 1240-1247, Jun. 2001.

Analyzing and Comparing the Parsing Techniques of Asynchronous Message

Mr.P.Krishna Sankar,
Assistant Professor,

Department of Computer science and Engineering,
Dr.Mahalingam College of Engineering &
Technology, Pollachi – 642 003, Tamilnadu, India.
pksankar@gmail.com

Ms.N.P.Shangaranarayane
Student,

Department of Computer science and Engineering,
Angel College of Engineering and Technology,
Tirupur-641 665, Tamilnadu, India.
npsnarayane@gmail.com

Abstract— Java API for XML Processing (JAXP) provided two methods for processing XML: Document Object Model (DOM) and Simple API for XML (SAX). The idea is to parse the whole document and construct a complete document tree in memory before it returns control to the client. This cannot be achieved through either by DOM nor by SAX. So StAX is introduced to achieve the idea. StAX does not suffer from the drawbacks faced while using DOM and SAX. A parser is a computer program or a component of a program that analyses the grammatical structure of an input with respect to a given formal grammar in a process known as parsing. Typically, a parser transforms some input text into a data structure that can be processed easily, e.g. for semantic checking, code generation or to help understanding the input. Such data structure usually captures the implied hierarchy of the input and forms a tree or even a graph. XML document as general tree structure and processing task as the extension from the parallel tree traversal algorithm for the classic discrete optimization problems. Unlike the Simple API for XML (SAX), StAX offers an API for writing XML documents. To be precise, it offers two APIs: a low-level, cursor-based API (XMLStreamWriter), and a higher-level, event-based API (XMLEventWriter). While the cursor-based API is best used in data binding scenarios (for example, creating a document from application data), the event-based API is typically used in pipelining scenarios where a new document is constructed from the data of input documents.

Keywords— DOM, SAX, StAX, API, XML

I. INTRODUCTION

XML stands for the Extensible Markup Language. It is a Markup language for documents, Nowadays XML is a tool to develop and likely to become a much more common tool for sharing data and store. XML can communicate structured information to other users [1]. In other words, if a group of users agree to implement the same kinds of tags to describe a certain kind of information, XML applications can assist these users in communicating their information in a more robust and efficient manner. XML can make it easier to exchange information between cooperating entities. XML technique can be categorized by four factors Strength of XML, XML Parser, XML Goals and Types of XML Parsers [5]. XML parsing is a core operation performed on an XML document for it to be accessed and manipulated. This operation is known to cause performance

bottlenecks in applications and systems that process large volumes of XML data [3].

XML processing can incur significant run-time overhead in XML-based infra structural middleware such as Web service application servers [4]. XML document is the general tree structure and the XML processing task as the extension from the parallel tree traversal algorithm for the classic discrete optimization problems. Analyse the standard parsing techniques like JDOM, SAX and STAX, based on that efficiency of different parsers is computed.

Java API for XML Processing (JAXP) provided two methods for processing XML - the Document Object Model (DOM) method, which uses a standard object model to represent XML documents, and the Simple API for XML (SAX) method, which uses application-supplied event handlers to process XML. Processing several XML documents simultaneously can be a significant challenge [2]. We are using the Java to develop the parser like JDOM, SAX, STAX and open source software. SAX parsers, for example, deliver the parsing events through callbacks to the client application. Because the SAX parser controls this process, the client application does not really have a chance to synchronize the different input sources. Therefore, programmers usually resort to the DOM parser when it comes to multi-document processing. However, the penalty here is excessive resource usage; the node trees of all input documents must completely reside in memory.

II. PARSING ANALYSIS

A parser is a computer program or a component of a program that analyses the grammatical structure of an input with respect to a given formal grammar in a process known as parsing. Typically, a parser transforms some input text into a data structure that can be processed easily, e.g. for semantic checking, code generation or to help understanding the input.

A. JDOM

JDOM is a tree-based API for processing XML documents with Java that threw out DOM's limitations and assumptions and started from scratch. It is designed purely for XML, purely for Java, and with no concern for backwards compatibility with earlier, similar APIs. JDOM is written in and for Java. It consistently uses the Java

coding conventions and the class library. It is thus much cleaner and much simpler than DOM. Most developers find JDOM to be far more intuitive and easy to use than DOM. It's not that JDOM will enable you to do anything you can't do with DOM. However, writing the same program with JDOM will normally take you less time and have fewer bugs when finished, simply because of the greater intuitiveness of the API. In many ways, JDOM is to DOM as Java is to C++, a much improved, incompatible replacement for the earlier more complex technology. JDOM is an open source, tree-based, pure Java API for parsing, creating, manipulating, and serializing XML documents. JDOM was invented by Brett McLaughlin and Jason Hunter in the spring of 2000.

JDOM can build a new XML tree in memory. Data for the tree can come from a non-XML source like a database, from literals in the Java program, or from calculations as in many of the Fibonacci number examples in this book. When creating new XML documents from scratch (rather than reading them from a parser), JDOM checks all the data for well-formed. For example, unlike many DOM implementations, JDOM does not allow programs to create comments whose data includes the double hyphen -- or elements and attributes whose namespace mapping conflict in impossible ways.

Once a document has been loaded into memory, whether by creating it from scratch or by parsing it from a stream, JDOM can modify the document. A JDOM tree is fully read-write. All parts of the tree can be moved, deleted, and added to, subject to the usual restrictions of XML.

B. SAX

SAX stands for Simple API for XML. SAX parsing is unidirectional; previously parsed data cannot be re-read without starting the parsing operation again. The SAX standard currently is at version 2.0. It is used to read data from a XML document. A parser that uses SAX parses the XML serially. The API is event driven and these events are fired when the XML features are encountered. XML parsing is unidirectional. Memory used by a SAX parser is relatively low. Due to the event nature of SAX, the parsing is faster of an XML document. SAX usually follows Push-based parsing, in which case, the Parser will scan the XML Document from top to bottom and whenever it finds some node (like start node, end node, text-node etc.) it will push notifications to the Application in the form of Events. So, SAX is basically a sequential, event-based parser. SAX is a callback implementation. As it iterates over each fundamental unit of XML, is that as it reads each unit of XML, it creates an event that the host program can use. This allows the application to ignore the bits it doesn't care about, and just keep or use what is needed. SAX is often used in certain high-performance applications or areas where the size of the XML might exceed the memory available to the running program. In mainstream languages, event-based interfaces are usually implemented using callback functions, a style familiar in graphical user interface (GUI) programming and the like. In object-oriented languages, callbacks are usually registered methods

for an object, using polymorphism to match the method name to the handler code, and using encapsulation to manage state in the handler between callbacks. This overall model of event-based programming is known as a push model and has a reputation for being difficult for many programmers to master. Most models that are considered easier to program, however, require random access to the document, and thus can lead to inefficiencies, so SAX has the reputation for being the most efficient standard way to process XML, if far from the easiest.

C. StAX

Streaming API for XML (StAX) is an application programming interface (API) to read and write XML documents, originating from the Java programming language community. Traditionally, XML APIs are either:

- Tree based - the entire document is read into memory as a tree structure for random access by the calling application
- Event based - the application registers to receive events as entities are encountered within the source document.

Streaming APIs for XML (StAX) which is a standardized Java based API for pull-parsing XML. StAX has two basic functions: to allow users to read and write XML as efficiently as possible and be easy to use (cursor API), and be easy to extend and allow for easy pipelining (event iterator API). Pull parsing differs from the traditional SAX based iteration and DOM based tree model, in that it is optimized for speed and performance. StAX is often referred to as "pull parsing." The developer uses a simple iterator based API to "pull" the next XML construct in the document. However, the common streaming APIs like SAX are all push APIs. They feed the content of the document to the application as soon as they see it, whether the application is ready to receive that data or not. SAX and XNI are fast and efficient, but the patterns they require programmers to adopt are unfamiliar and uncomfortable to many developers.

Pull APIs are a more comfortable alternative for streaming processing of XML. A pull API is based around the more familiar iterator design pattern rather than the less well-known observer design pattern. In a pull API, the client program asks the parser for the next piece of information rather than the parser telling the client program when the next datum is available. In a pull API the client program drives the parser. In a push API the parser drives the client.

Reading with the StAX is by XMLStreamReader .It is the key interface in StAX. This interface represents a cursor that's moved across an XML document from beginning to end. At any given time, this cursor points at one thing: a text node, a start-tag, a comment, the beginning of the document, etc. The cursor always moves forward, never backward and normally only moves one item at a time.

There are a few ways to filter the event stream; of course, you could use a stack of if-else statements instead of the switch, but almost all StAX programs will feature an

event loop something like this one. This is probably my only major criticism of StAX. Integer type codes and big switch statements are relics of procedural thinking. Object oriented programs should be based around classes, inheritance hierarchies, and polymorphism instead.

StAX is a fast, potentially extremely fast, straightforward, memory-thrifty way to loading data from an XML document the structure of which is well known in advance. State management is much simpler in StAX than in SAX, so if you find that the SAX logic is just getting way too complex to follow or debug, then StAX is well worth exploring. A few features such as validation, schema support, and entity resolution are either not available or are not functional in the current reference implementation, but these should soon be available in independent implementations [6]. StAX will be a very useful addition to any Java developer's XML toolkit.

III.COMPARISION

Java API for XML Processing (JAXP) provided two methods for processing XML -- the Document Object Model (DOM) method, which uses a standard object model to represent XML documents, and the Simple API for XML (SAX) method, which uses application-supplied event handlers to process XML. Processing several XML documents simultaneously can be a significant challenge. SAX parsers, for example, deliver the parsing events through callbacks to the client application. Because the SAX parser controls this process, the client application does not really have a chance to synchronize the different input sources. Therefore, programmers usually resort to the DOM parser when it comes to multi-document processing. However, the penalty here is excessive resource usage; the node trees of all input documents must completely reside in memory. In each step of the parsing Java object model is to be performed (Fig.1).

The screening or classification of XML documents is a common problem, especially in XML middleware. Routing XML documents to specific processors may require analysis of both the document type and the document content. The problem here is obtaining the required information from the document with the least possible overhead. Traditional parsers such as DOM or SAX are not well suited to this task. In the Fig.2 the XML document is converting to the Java Object model.

DOM, for example, parses the whole document and constructs a complete document tree in memory before it returns control to the client. Even DOM parsers that employ deferred node expansion, and thus are able to parse a document partially, have high resource demands because the document tree must be at least partially constructed in memory. This is simply not acceptable for screening purposes.

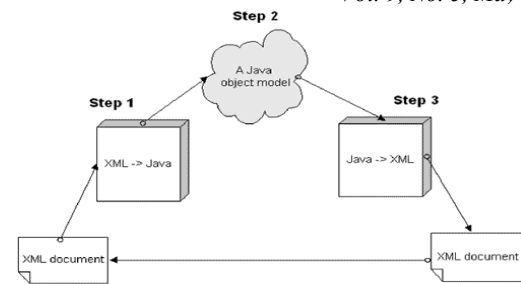


FIG.1. PARSING OF XML DOCUMENT

Like DOM, SAX parsers control the complete parsing process. By default, a SAX parser starts parsing at the beginning of a document and continues until the end. Client event handlers are informed through callbacks about the events during this parsing process. To avoid unnecessary overhead during document screening, such an event handler may want to stop the parsing process once it has gathered the required information. A common technique for achieving this in SAX is throwing an exception. This will cause SAX to stop the parsing process.

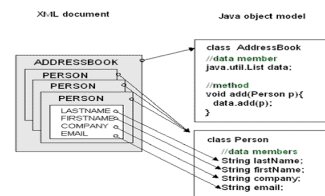


FIG.2. CONVERT THE XML INTO JAVA OBJECT MODEL

The information gathered by the event handler must be encoded in an error message that's wrapped in an exception object and posted to the parser's client. A special error handler in the client receives this exception and must parse the parser's error message to retrieve the required information. This may be a solution to the screening problem, but it's a complicated one. SAX parsers, for example, deliver the parsing events through callbacks to the client application. Because the SAX parser controls this process, the client application does not really have a chance to synchronize the different input sources. Therefore, programmers usually resort to the DOM parser when it comes to multi-document processing. However, the penalty here is excessive resource usage; the node trees of all input documents must completely reside in memory.

StAX does not suffer from above drawbacks. As its name indicates, it is targeted at streaming applications such as the merging of two documents. The following example shows how this is done. Assume that you want to merge two documents containing lists of products.

Streaming API for XML (StAX) completely changes this. Unlike the Simple API for XML (SAX), StAX offers an API for writing XML documents. To be precise, it offers two APIs: a low-level, cursor-based API (XMLStreamWriter), and a higher-level, event-based API (XMLEventWriter). While the cursor-based API is best used in data binding scenarios (for example, creating a document from application data), the event-based API is

typically used in pipelining scenarios where a new document is constructed from the data of input documents. The cursor-based API offers a variety of specific methods for creating the various elements of the XML information set, such as elements, attributes, processing instructions, data type declarations, and character content. These methods take care of many formatting issues. For example, the method `writeCharacters()` automatically escapes characters like the less than sign (<), the greater than sign (>), and the ampersand (&). And the method `writeEndElement()` automatically closes all open structures. So it does not matter if the last call to `writeEndElement()` in the example is commented out or not.

StAX can even generate namespace prefixes for namespaces that have not been formally declared. `javax.xml.stream.isPrefixDefaulting` has been set to true for the output factory. If this property has been set to false, you must explicitly declare each namespace prefix and each namespace using the methods `setPrefix()` and `writeNamespace()`. Among the DOM and SAX widely used methods, StAX provides the parsing efficiency and making developer comfort. As StAX name indicates, it is targeted at streaming applications such as the merging of two documents and exchange information between cooperating entities.

StAX allows an application to process multiple XML sources simultaneously. For example: when one document includes or imports another document, the application can process the imported document while processing the original document. This use case is common when the application is reading documents such as XML Schemas or WSDL documents [4]. StAX has two basic functions: To allow users to read and write XML as efficiently as possible and be easy to use (cursorAPI), and be easy to extend and allow for easy pipelining.

This approach of XML processing gives more control to the client application than to the parser, enabling faster and memory-efficient processing. This is becoming a standard across different domains of XML processing. For example, Apache Axis2, one of the prominent SOAP processing engines, improved its performance four times, on average, over its predecessor by using a StAX-based XML processing model called Axiom. Axiom is more memory-efficient and preferment than the existing object models available today due to the usage of StAX as its XML parsing technology.

TABLE I COMPARING THE JDOM, SAX, STAX

Parser APIs	Advantages	Disadvantages
DOM	<ul style="list-style-type: none">Rich set of APIsEasy navigationEntire tree loaded into memory, random access to XML document	<ul style="list-style-type: none">XML document must be parsed at one timeExpensive to load entire tree into memoryGeneric DOM node not ideal for object-type binding

SAX	<ul style="list-style-type: none">Entire document is not loaded into memory, resulting in low memory consumptionAllows registration of multiple content Handlers.	<ul style="list-style-type: none">No built-in document navigation supportNo random access to XML DocumentNo support for modifying XML in place.No support for namespace scoping.
StAX	<ul style="list-style-type: none">Contains two parsing models, for ease of performance.Application controls parsing, easily supporting multiple inputsPowerful filtering capabilities provide efficient data retrieval	<ul style="list-style-type: none">No built-in document navigation supportNo random access to XML documentNo support for modifying XML in placeStill in an immature state

IV.RESULT

Based on time taken to parsing of the xml content with JDOM, SAX and StAX techniques get data.

TABLE II TIME TAKEN QUADCORE PROCESSOR TO PARSE AN XML

Nodes	Time taken (nanoseconds)		
	JDOM	SAX	StAX
1	0.0543514280	0.0302972850	0.0229867110
2	0.0551824870	0.0305735140	0.0230995120
3	0.0552194060	0.0308396360	0.0231593400
4	0.0552609610	0.0310450710	0.0232124300
5	0.0552998820	0.0312222530	0.0232675710
6	0.0553378660	0.0314174210	0.0233201160
7	0.0553779660	0.0315977420	0.0233766030
8	0.0554156950	0.0317844120	0.0234354000
9	0.0554545450	0.0319676010	0.0234957610
10	0.0554942100	0.0321462610	0.0235535530
11	0.0555336980	0.0323332980	0.0236119220
12	0.0555771380	0.0325268310	0.0236688970
13	0.0556225780	0.0327048640	0.0237263470
14	0.0556670760	0.0328829090	0.0237855310
15	0.0557098490	0.0330663200	0.0238462250
16	0.0557561020	0.0332445070	0.0239056370
17	0.0557997820	0.0334221430	0.0239639500
18	0.0558468040	0.0336052810	0.0240300790
19	0.0558944750	0.0337868240	0.0240976250
20	0.0559429890	0.0339688360	0.0241573290
21	0.0559919680	0.0341568860	0.0242198540

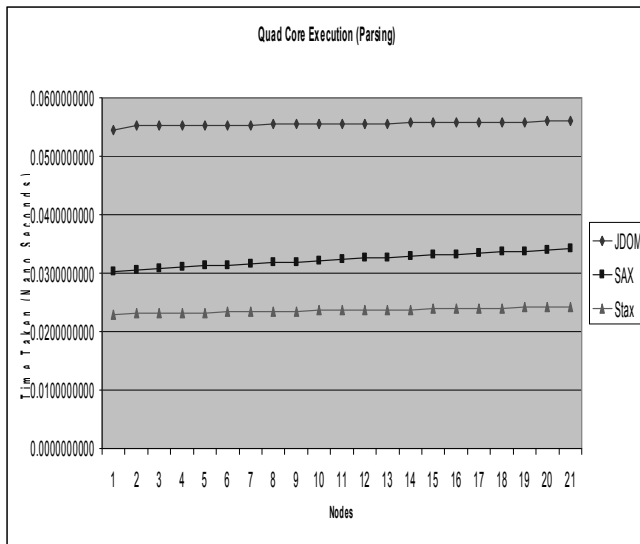


FIG.3. TIME TAKEN BY QUADCORE PROCESSOR TO PARSE AN XML

TABLE II TIME TAKEN PENTIUM D PROCESSOR TO PARSE AN XML

Nodes	Time taken (nanoseconds)		
	JDOM	SAX	StAX
1	0.5790705250	0.1829763280	0.0863534340
2	0.6085354430	0.1833428550	0.0865333450
3	0.6086161790	0.1836504360	0.0866400620
4	0.6086890940	0.1839482390	0.0867403540
5	0.6087569790	0.1842457630	0.0868412050
6	0.6088259820	0.1859532420	0.0869440110
7	0.6088958240	0.1862731160	0.0870443030
8	0.6089712520	0.1865784620	0.0871476690
9	0.6090441670	0.1868706780	0.0872543860
10	0.6091218300	0.1871617760	0.0873608240
11	0.6092008900	0.1874640490	0.0875563800
12	0.6092768780	0.1877724680	0.0876628180
13	0.6093567760	0.1880747410	0.0877661830
14	0.6094341600	0.1883831600	0.0878751350
15	0.6095171320	0.1886907410	0.0879947040
16	0.6095973100	0.1891279480	0.0881047730
17	0.6096822370	0.1894360870	0.0882223860
18	0.6097646490	0.1897319350	0.0883324560
19	0.6098479000	0.1900311350	0.0884402910
20	0.6099308710	0.1903364810	0.0885922660
21	0.6100155190	0.1922856120	0.0887062470

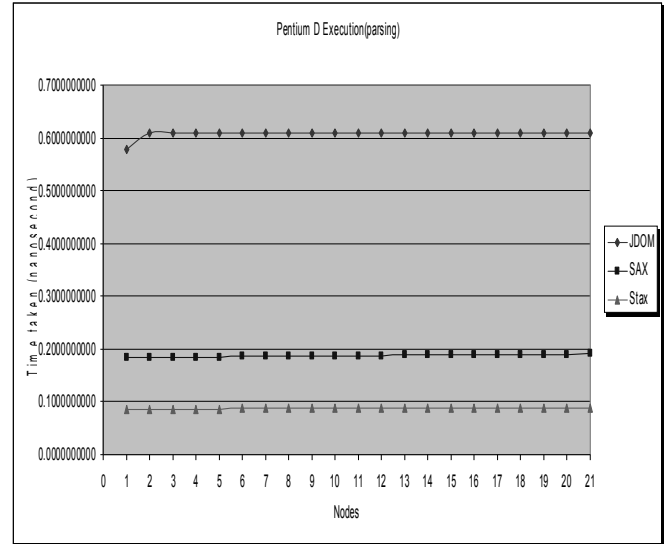


FIG.4. TIME TAKEN BY PENTIUM D PROCESSOR TO PARSE AN XML

From the above graph in Fig 3 and Fig 4, StAX takes minimum time than other parsers JDOM and SAX.

V.CONCLUSIONS

Java API for XML Processing (JAXP) which processing XML documents by using, the Document Object Model (DOM) method, the Simple API for XML (SAX) method, and Streaming API for XML (StAX) method are used commonly. As StAX name indicates, it is targeted at streaming applications such as the merging of two documents and exchange information between cooperating entities. StAX allows an application to process multiple XML sources simultaneously. Among the DOM and SAX widely used methods, StAX provides the parsing efficiency and making developer comfort.

REFERENCES

- [1] Rami Alnaqeb, Fahad H.Alshammari ,Zaidan.M.A., Zaidan.A.A., Zaidan.B.B., Zubaidah M.Hazza (2010) "An Overview: Extensible Markup Language Technology", in Journal of computing, Volume 2, Issue 6, June 2010.
- [2] Nayak, Richi and Witt, Rebecca and Tonev, Anton (2002) "Data mining and XML documents", In Proceedings International Conference on Internet Computing, IC'2002 3, pages pp. 660-666.
- [3] Guy Lapalme (2010) "Exploring and Extracting Nodes from Large XML Files" available at <http://www.iro.umontreal.ca/~lapalme/ExamineXML>.
- [4] Wei Zhang and Robert A. van Engelen (2009) "An Adaptive XML Parser for Developing High-Performance Web Services", in proceedings of the International Symposium on Web Services (ISWS).
- [5] Antoniu Nicula, Doina Zmaranda and Codruța Vancea (2010) "Issues on efficiency of XML parsers", <http://www.rpbouret.com/xml/XMLData Binding>.
- [6] Morris Matsa, Eric Perkins, Abraham Heifets, Margaret Gaitatzes Kostoulas, Daniel Silva, Noah Mendelsohn, Michelle Leger (2007) "A High Performance Interpretive Approach to Schema Directed Parsing", in International World Wide Web Conference (IW3C2) paper.
- [7] Toshiro Takase, Hisashi Miyashita, and Toyotaro Suzumura, Michiaki Tatsubori, (2005) "An Adaptive, Fast, and Safe XML Parser Based on Byte Sequences Memorization" in International World Wide Web Conference (IW3C2) may 2005.

AUTHOR PROFILE

1. Mr.P.Krishna Sankar, M.E., Working as Assistant Professor, Dr. Mahalingam College of Technology, Pollachi, Coimbatore(Dt), Tamilnadu (State), India. Earlier, worked in IBM – India Software Lab, Bangalore as Software Engineer for a year. Completed Master of Engineering in Computer Science and Engineering in PSG College of Technology, Coimbatore (Dt), Tamilnadu (State), India and Bachelor of Engineering in Computer Science and Engineering in K.S.Rangasamy College of Technology, Tiruchengode, Namakkal (Dt), Tamilnadu (State), India. Have published 3 National Conference papers and performed 8 Technical Paper Presentations.
2. Ms.N.P.Shangaranarayane, Student pursuing Bachelor of Engineering in Computer Science and Engineering in Angel College of Engineering and Technology, Tirupur (Dt), Tamilnadu (State), India. Have published 1 National Conference paper and performed 4 Technical Paper Presentations.



Analysis on Differential Router Buffer Size towards Network Congestion

A Simulation-based

Haniza N., Zulkiflee M., Abdul S. Shibgatullah, Shahrin S.

Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

haniza@utem.edu.my, zulkiflee@utem.edu.my, samad@utem.edu.my, shahrinsahib@utem.edu.my

Abstract—Network resources are shared amongst a large number of users. Improper managing network traffic leads to congestion problem that degrades a network performance. It happens when the traffic exceeds the network capacity. In this research, we plan to observe the value of buffer size that contributes to network congestion. A simulation study by using OPNET Modeler 14.5 is conducted to achieve the purpose. A simple dumb-bell topology is used to observe several parameter such as number of packet dropped, retransmission count, end-to-end TCP delay, queuing delay and link utilization. The results show that the determination of buffer size based on Bandwidth-Delay Product (BDP) is still applicable for up to 500 users before network start to be congested. The symptom of near-congestion situation also being discussed corresponds to simulation results. Therefore, the buffer size needs to be determined to optimize the network performance based on our network topology. In future, the extension study will be carried out to investigate the effect of other buffer size models such as Stanford Model and Tiny Buffer Model. In addition, the buffer size has to be determined for wireless environment later on.

Keywords – OPNET, network congestion, bandwidth delay product, buffer size

I. INTRODUCTION

Router plays an important role in switching packet over a public network. A storage element called as buffer is responsible to manage transient packets in a way of determining its next path to be taken and deciding when packets suppose being injected into network. Several studies [1-3] have agreed that the single biggest contributor to the uncertainty of Internet is coming from misbehavior of router buffer per se. It introduces some queuing delay and delay-variance between flow transitions. In some cases, packets are potential to be lost whenever buffer is overflow. Oppositely, it is wasteful and ineffective when buffer is underutilized. As a result, it shows some degradation in the expected throughput rate.

The main factor to increase the network performance is to seize the optimal size of router buffer. Currently, it is set either a default value specified by the manufacturer or it is determined by the well known “Bandwidth-Delay Product”

(BDP) principal that has been invented by [4]. This rule is aiming to keep a congested link as busy as possible and maximize the throughput while packets in buffer were kept busy by the outgoing link. The BDP buffer size is defined as an equal to the product of available data link’s capacity and its end-to-end delay at a bottleneck link. The end-to-end delay can be measured by Round-Trip Time (RTT) as presented in Equation (1). The number of outstanding packets (in-flight or unacknowledged) should not exceeds from TCP flow’s share of BDP value to avoid from packet drop[5].

$$BDP \text{ (bits)} = \text{Available Bandwidth (bits/sec)} \times RTT \text{ (sec)} \quad (1)$$

In ideal case, the maximum packets carrying in a potential bottleneck link can be gain from a measurement of BDP_UB where there is no competing traffic. The BDP_UB or Upper Bound is given in Equation (2) as stated below:

$$BDP_UB \text{ (bits)} = \text{Total Bandwidth (bits/sec)} \times RTT \text{ (sec)} \quad (2)$$

When applied in the context of the TCP protocol, the size of window sliding should be large enough to ensure that enough in-flight packets can put in congested link. To control the window size, TCP Congestion Avoidance uses Additive Increase Multiple Decrease (AIMD) to probe the current available bandwidth and react against overflow buffer. The optimal congestion window size is expected to be equal to BDP value; otherwise packet will start to queue and then drop when it “overshoots”.

Today, several studies have been conducted to argue the realistic of BDP such as Small buffer which also known as Stanford Model [6] and Tiny Buffer Model [7]. They keep try to reduce number of packets in buffer without loss in performance. Larger buffers have a bad tradeoff where it increases queuing delay, increase round-trip time, and reduces load and drop probability compared to small buffers which have higher drop probability [8]. However, applications able to protect against packet drop rather than recapture lost time.

The goal of this paper is to study the effectiveness of BDP on a simple network topology. This will be demonstrated on a group of users from a range of 5 until 1000 users. A simulation study is carried out with OPNET Modeler 14.5 [9].

The rest of the paper is organized as follows. Section II reviews the term of congestion from several aspects and briefly explain about a well known buffer sizing model, BDP. Section III describes the network model and evaluation metrics for the simulation. In Section IV, we analyses simulation results. Section V concludes the present paper and discusses some possible extensions of our work.

II. BACKGROUND STUDY

A. Congestion

In [10] stated that network congestion was related to the buffer space availability. For normal data transmission, the number of packet sent is proportional to the number of the packets delivered at destination. When it reaches at saturation point and packets still being injected to network, a phenomenon called as Congestion Collapse will be occurred. In this situation, the space buffer considers limited and fully occupied. Thus, the incoming packets need to be dropped. As a result, a network performance has been degraded.

Most previous studies [11-13] emphasized that the key of congestion in wired network is from network resources limitation. This limitation is including the characteristics of buffer, link bandwidth, processor times, servers, and forth. In a simple Mathematical definition, congestion occurred once there are more demands exceed the available network resources as represented by Equation (3).

$$\sum Demand > Available Resources \quad (3)$$

In [13], the congestion problem has been widely defined from different perspectives including Queue Theory, Networking Theory, Network Operator and also Economic aspect. However, it still emphasizes on buffer-oriented activity and capability to handle unexpected incoming packets behavior. For instance, the access rate exceeds the service rate at intermediate nodes.

B. Rule of thumb

Most routers in the backbone of the Internet have a Bandwidth-Delay Product (BDP) of buffering for each link. This rule has been concluded based on an experimental of a small number of long-lived TCP flows (eight TCP connections) on a 40 Mbps link. The selection of TCP flows has been proved by [14, 15] that more than 90 % of network traffics is TCP-based. Meanwhile, the value of BDP that more than 10^5 bits (12500 bytes) is applicable for Long-Fat Network (LFN). In this case it refers to Satellite Network [16].

III. METHODOLOGY

In this study, a proper methodology has been designed to get an expected output. This can be referred to the following work flow depicts in Figure 1.

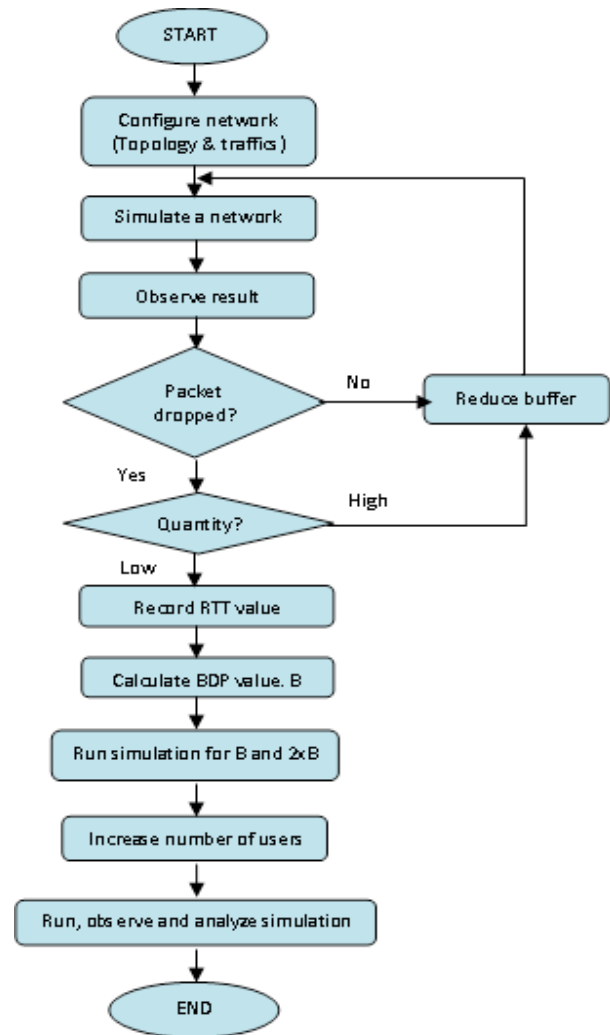


Figure 1: Methodology to be used

The first step demands for defining the value of the Round-Trip Time (RTT). This value can be set based on a normal data transmission where there no packets drop yet. To achieve it, the network needs to be configured based by using a default setting that available in simulation tool. Then, the memory size at router need to be adjusted until last configuration where there a small number of packets dropped appear. Once RTT has successfully estimated, a current buffer size will be recorded and then need to adjusted base on BDP model.

The next step is to compare the effect of different buffer size as mentioned previously in Section I. There are two scenarios created to represent Scenario 1 (Small Buffer **B**) and scenario 2 (Large Buffer **2xB**). Both scenarios will be tested for a different range of users from 5 to 1000. Several parameters will be observed and then analyzed more detail later. This simulation will be run for 900 seconds.

IV. EXPERIMENTAL APPROACH

A. Network Environment Setup

In this section, a simple network topology which also known as dumb-bell topology was designed as illustrated in Figure 2. This topology is a typical model used by researcher to study congestion issues as stated in [17]. The network consists of three servers, LAN users, two intermediate routers and links interconnecting between them. For both links between servers/LAN users, the data rate is given as 100 Mbps. Meanwhile, routers are connected using Point-to-point Protocol (PPP) 1.544 Mbps.

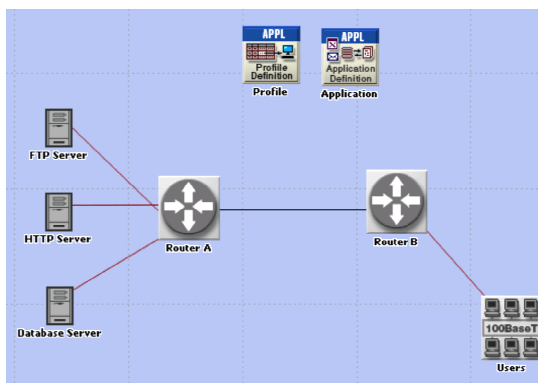


Figure 2. Proposed system network

For application configuration, TCP-based services such as File Transfer Protocol (FTP), Database and web browsing traffic (HTTP) were defined. Table 1 shows the traffic definition that used in our simulation.

TABLE 1 TRAFFICS DEFINITION FOR SIMULATION

Services	Description	Value
FTP	Command Mix (Get/Total) :	50%
	Inter-Request Time (seconds) :	360
	File size (bytes):	1000
Database	Transaction Mix (Queries/ Total Transaction) :	100%
	Transaction Interarrival Time (seconds) :	12
	Transaction Size (bytes) :	32768
	HTTP Specification :	HTTP 1.1
HTTP	Page Interarrival Time (seconds) :	10

B. Evaluation Metrics

In this study, the behavior of the packet once passing throughout Router B was observed. This study assumed that router maintains a single FIFO queue, and drop packets from the tail when the queue is full. This action is known as Drop-tail which is the most widely deployed scheme today. We collect some useful information such as number of packet dropped, retransmission count, end-to-end TCP delay, queuing delay and link utilization. This selection based on possible output to represent a possible picture of congestion phenomenon in the network topology.

V. SIMULATION RESULTS & ANALYSIS

In this section, simulation result for the impact of changing buffer sizes on network performance was presented. Simulations were run for Bandwidth-Delay Product (BDP) model. Based on Equation (1), we used two values of buffer sizes which are **B = 2000 bytes**, referred as the “small buffer” and another is given as **B = 4000 bytes**, referred as the “large buffer”. This BDP values were calculated to show the differences buffer space availability towards network congestion.

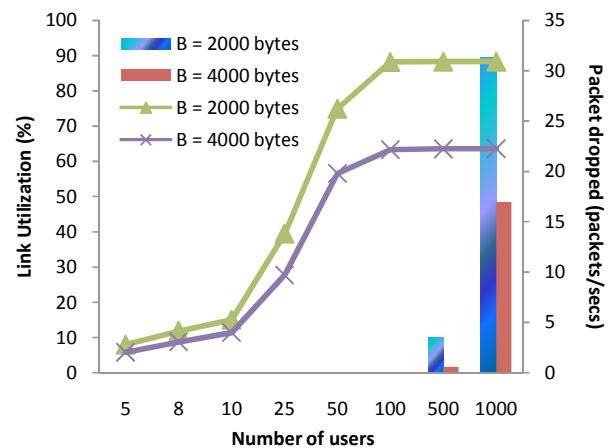


Figure 3: The influence buffer size to link utilization and packet drop

Figure 3 shows the influence buffer size to link utilization and packet drop when the number of users N is changed. To be clear, the line graph represents link utilization meanwhile the bar chart represents packet drop activity. For both graphs, it can be seen that “small buffer” always obtained high link utilization and high packet drop compared to “large buffer”. To analyze this simulation result, we divide users into three grouping: Group A, Group B and Group C as shown in Table 2.

TABLE 2. USERS GROUP

	Group A	Group B	Group C
Users	5-10	11-100	101-1000

For link utilization, we found that small number of user in Group A for only occupy backbone link less than 20%. Meanwhile Group B which has medium number of users keeps increases its link usage until 60-90% from the available link. However, the link utilization for Group C has remained at almost 60% (large buffer) and 90% (small buffer). This link saturation caused by buffer space limitation in Router B for both cases when it considered as fully occupied. As a result, the incoming packets start to be dropped.

For the bar chart information, the packet discarded obviously in Group C particularly when users count more than 500. The higher packet dropped was slightly 30 packets/second for “small buffer” and slightly 15 packets/second for “large buffer”. It can be conclude that buffer space is still available and no packet drop when users is in Group A and Group B for BDP model.

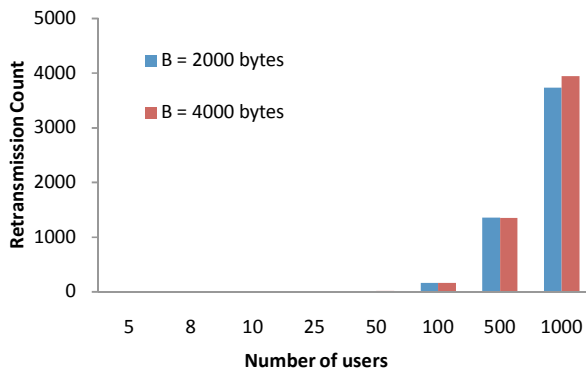


Figure 4: Packet retransmission

Figure 4 depicts the number of packet retransmission when the number of users N is changed. It can be seen that the retransmission activity has been detected started when the user reached 50 for “large buffer” and 100 for small buffer size. Based on TCP Congestion Control specification [18], each delivered packets must be acknowledged in time. If timeout or packets delay, sender will automatically do packet retransmission. By default, retransmission attempts are allowed not more than 3 times in sequence. If exceeds, the packet is assumed to be lost and then TCP Congestion Control mechanism will start to halve congestion window and reduce sending rate.

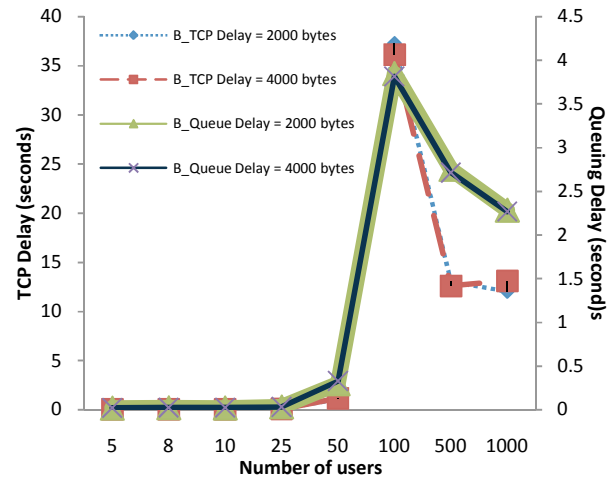


Figure 5: End-to-end TCP delay and Queuing Delay for different users

Figure 5 shows the End-to-end TCP delay and Queuing delay when the number of users N is changed. For both delays, it kept to increase rapidly when user between a range of 50 to 100. However, these delays start to drop when the link between routers started to be saturated. This action result from TCP congestion control that applies rate adaptation once network congested.

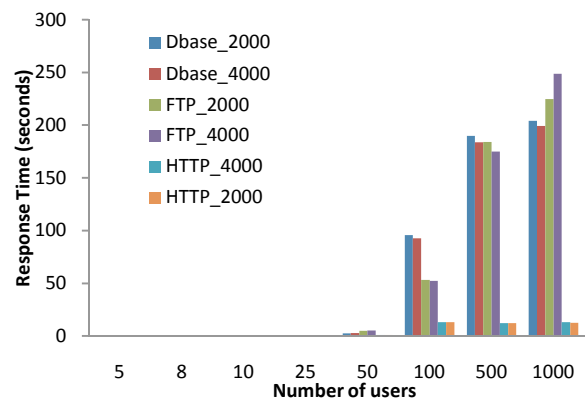


Figure 6: The influence of buffer size on the Application Response time

Figure 6 illustrates the influence of the buffer size on the applications response time when the number of users N is changed. For both buffer sizes, it can be seen that FTP and Database applications has higher response time compared HTTP services.

In summary, the determination of buffer size based on the Bandwidth-Delay product (BDP) gives a value of small buffer ($B = 2000$ bytes) and large buffer ($B = 4000$ bytes) to be used in understanding of their effects on network performance. By taking consideration on the influence of the growth of users in network, the packet behavior has

been observed correspond to the availability of router buffer space such as link utilization, packet dropped, retransmission count, end-to-end TCP delay, queuing delay and application's response time.

From the simulation result discussed above, the buffer start to be congested when the user reach to 500. This assumption was based on situation where there are higher link utilization and higher packets dropped. The symptom of near-congestion situation can be observed from activities such as packets retransmission, end-to-end TCP delay, queuing delay and application response time. This symptom occurred when users are between 25 and 50.

VI. CONCLUSION

In this paper, the effect of router buffer size based on Bandwidth-Delay Product (BDP). Through a simulation, the value of small buffer is important element rather than large buffer in order to have better network performance. This also depends on number of users and applications running on a network. In the future, we plan to investigate the effect of other buffer size models such as Stanford Model and Tiny Model. Furthermore, the buffer size has to be determined for in wireless environment later on.

ACKNOWLEDGEMENT

The research presented in this paper is supported by Malaysian government scholarship and it was conducted in Faculty of Information and Communication Technology (FTMK) at Universiti Teknikal Malaysia Melaka, Malaysia.

REFERENCES

- [1] Wischik, D. and N. McKeown, *Part I: Buffer sizes for core routers*. ACM SIGCOMM Computer Communication Review, 2005. 35(3): p. 75-78.
- [2] Appenzeller, G., I. Keslassy, and N. McKeown, *Sizing router buffers*. ACM SIGCOMM Computer Communication Review, 2004. 34(4): p. 281-292.
- [3] Welzl, M., *Network congestion control*. 2005: Wiley Online Library.
- [4] Villamizar, C. and C. Song, *High performance TCP in ANSNET*. ACM SIGCOMM Computer Communication Review, 1994. 24(5): p. 45-60.
- [5] Chen, K., et al., *Understanding bandwidth-delay product in mobile ad hoc networks*. Computer Communications, 2004. 27(10): p. 923-934.
- [6] Dhamdhere, A. and C. Dovrolis, *Open issues in router buffer sizing*. ACM SIGCOMM Computer Communication Review, 2006. 36(1): p. 87-92.
- [7] Raina, G. and D. Wischik, *Buffer sizes for large multiplexers: TCP queueing theory and instability analysis*. 2005: IEEE.
- [8] Tomioka, T., G. Hasegawa, and M. Murata, *Router buffer re-sizing for short-lived TCP flows*.
- [9] *OPNET Modeler-version 14.5*.
- [10] Nagle, J., *Congestion control in IP/TCP internetworks*. ACM SIGCOMM Computer Communication Review, 1984. 14(4): p. 11-17.
- [11] Jain, R., *Congestion control in computer networks: issues and trends*. Network, IEEE, 1990. 4(3): p. 24-30.
- [12] Keshav, S., *Congestion control in computer networks*. 1991: Univ. of California.
- [13] Bauer, S., D. Clark, and W. Lehr, *The evolution of internet congestion*. 2009.
- [14] Low, S.H., F. Paganini, and J.C. Doyle, *Internet congestion control*. Control Systems Magazine, IEEE, 2002. 22(1): p. 28-43.
- [15] Li, T. and D.J. Leith, *Buffer sizing for TCP flows in 802.11 e WLANs*. Communications Letters, IEEE, 2008. 12(3): p. 216-218.
- [16] Van Jacobson, R.B. and D. Borman, *TCP extensions for high performance*. 1992.
- [17] Floyd, S. and E. Kohler, *Internet research needs better models*. ACM SIGCOMM Computer Communication Review, 2003. 33(1): p. 29-34.
- [18] Allman, M., V. Paxson, and W. Stevens, *TCP congestion control*. 1999.

AUTHORS PROFILES

Haniza Nahar, a Senior Lecturer at University of Technical Malaysia Melaka (UTeM). She earned MSc. in ICT for Engineers (Distinction) from Coventry University, UK and BEng. in Telecommunication from University Malaysia. She used to be an Engineer and has been qualified for CFOT and IPv6 Software Engineer. Her postgraduate dissertation has been awarded as the **Best Project Prize**.

Zulkiflee Muslim, a Senior Lecturer at University of Technical Malaysia Melaka (UTeM). He earned MSc. in Data Communication and Software from University of Birmingham City, UK and BSc. in Computer Science from University of Technology Malaysia. He has professional certifications: CCNA, CCAI, CFOT and IPv6 Network Engineer Certified.

Dr. Abdul Samad Shibghatullah, a Senior Lecturer at University of Technical Malaysia Melaka (UTeM). He earned MSc Computer Science from Universiti Teknologi Malaysia (UTM) and B.Acc from Universiti Kebangsaan Malaysia (UKM). His areas of research include Scheduling and Agent Technology.

Prof. Dr. Shahrin bin Sahib @ Sahibuddin is a Dean of Faculty of Information and Communication Technology, UTeM. He earned PhD in Parallel Processing from Sheffield, UK; MSc Eng System Software and B Sc Eng Computer Systems from Purdue University. His areas of research include Network, System, Security, Network Admin and Design.

Mobility Assisted Solutions for Well-known Attacks in Mobile Wireless Sensor Network

Abu Saleh Md. Tayeen

Department of Computer Science
American International University- Bangladesh (AIUB)
Dhaka, Bangladesh
tayeen@aiub.edu

A.F.M. Sultanul Kabir

Department of Computer Science
American International University- Bangladesh (AIUB)
Dhaka, Bangladesh
afmk@kth.se

Razib Hayat Khan

Department of Telematics
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
rkhan@item.ntnu.no

Abstract— Over the past few years the domain of wireless sensor networks applications is increasing widely. So security is becoming a major concern for WSN. These networks are generally deployed randomly and left unattended. These facts coupled together make it vulnerable to different dangerous attacks like node capture attack, node replication attack, wormhole attack, sinkhole attack etc. Several detection schemes and countermeasures have been proposed in the literature to defend against such attacks in static sensor networks. However these solutions rely on fixed locations of sensor nodes and thus do not work in mobile wireless sensor networks where sensor nodes are expected to have mobility nature. This paper provides summarization of typical attacks on mobile wireless sensor networks and survey about the literatures on few important countermeasures relevant to these attacks.

Keywords- *Mobility, Mobile Wireless Sensor Networks, Security, Mobile Nodes.*

I. INTRODUCTION

Wireless sensor network has become one of the most promising technologies over the past few years. A WSN (wireless sensor network) is a multi-hop wireless network consisting of large number of distributed autonomous devices using sensors to cooperatively perform a common task. Though the development of wireless sensor networks was originally motivated by military applications such as battle field surveillance, now WSN is used in wide range of potential applications including environment and habitat monitoring, object tracking, scientific observing and forecasting, industrial, medical, traffic control applications and etc.

WSNs are often deployed with no existing infrastructure and left unattended. Security becomes a critical challenge when sensor networks are used in hostile environments where they are exposed to various malicious attacks. For example, an adversary can easily monitor the entire network communications, capture legitimate sensor nodes (node capture attack) to acquire all the credential information stored therein and launch node replication attacks, wormhole attacks and sinkhole attacks. Most of the current schemes to defend against these attacks are only suitable for static WSN. That means

existing techniques assume that sensor nodes and the base station are stationary. However, there may be situations, such as battlefield environments, where the base station and possibly the sensors need to be mobile. The mobility of sensor nodes has a great influence on sensor network topology and thus raises many issues for security.

The goal of this paper is to provide a brief description of some dangerous attacks on mobile wireless sensor network and outline possible solutions for each attack. To the best of our knowledge there has not been a comprehensive study of attacks for mobile wireless sensor networks. In this paper we attempt to give such a comprehensive survey. The remainder of the paper is organized as follows. Section II describes about the node capture attack which represents the first step to further attacks. Section III deals with the node replication or clone attack that is considerably difficult to detect in sensor networks. Wormhole attacks are discussed in Section IV. Section V presents the sinkhole attack which can cause serious problem to the operations and services of sensor networks and conclusion are drawn in Section VI.

II. NODE CAPTURE ATTACK

Node capture attack [1] is one of the most important and challenging issues of wireless sensor network security. Due to the nature of operations, in most applications sensor nodes are likely to be placed in locations readily accessible to attackers. Such exposure allows the attackers to gain full control over some sensor nodes. As physical tamper proofing features are impractical because of low cost of the nodes, an attacker might capture sensor nodes, extract key information stored in nodes' memory, modify their programming or replace them with malicious nodes under the control of the adversary. Capturing nodes not only provides the adversary with an opportunity of unrestricted access to the entire wireless sensor network but also gives an effective way to influence the outcome of network protocols. This access into the network and control of sensor nodes can then be used as a basis for further attacks like Sybil attack [2] and Clone attack [3].

To protect wireless sensor networks against node capture attack, the event must be detected as early as possible. After detection the compromised node's id can be revoked from the network so that the maliciously modified node may not take part to the network operations in future. Although there are some solutions for node capture detection [4], two efficient and distributed detection schemes (SDD, CDD) has been proposed in [5]. The main focus of the distributed solutions is to use the mobility of sensor nodes to detect possible node captures. The simple observation of the solutions can be defined as follows. If node a has met with node b , that is, node a and b are in communication range and node a has heard a transmission from node b at a certain time t , and later it does not *re-meet* node b within a period λ , then node a can deduce that node b has possibly been captured.

A. SDD: Simple Distributed Detection

The event based Simple Distributed Detection protocol is initially proposed by Conti. In this protocol each node a is assigned the task to track a specific set T_a of other nodes. Whenever a gets into the communication range of any node $b \in T_a$, it sets the corresponding meeting time to the value of its internal clock and start the corresponding time-out period to λ seconds. If the nodes did not re-meet, (that is, the time out expires) node a triggers an alarm which is flooded to the network. This alarm is supported by the feature of revocation of node b .

B. CDD: Cooperative Distributed Detection

The Cooperative Distributed Detection protocol uses node cooperation in addition to mobility to greatly improve the performance of node capture detection. In this protocol two nodes a and b exchange information about the nodes in $T_a \cap T_b$ (that is nodes tracked by both a and b). Although the node cooperation requires more energy consumption (due to the message exchange), it allows to reduce the number of false positive alarms (nodes that are revoked even though they have not actually been captured), which is a desirable feature of node capture detection protocols.

C. Performance Analysis

A performance comparison of SSD and CDD protocols has been done through simulation against the Detection Time, False Positive and Energy Consumption parameters. In the simulation random way-point mobility model is used as node mobility pattern.

Detection Time: Detection time is the delay between actual node capture and detection. The shorter the detection time the better the protocol. The simulation results show that CDD is better than SDD for detecting node capture attack.

False Positive: False positive refers to the revocation done when the nodes are actually not being captured. A good node capture detection protocol is expected to reduce the number of false positives. Based on the simulation results CDD can decrease more false positives than SDD protocol.

Energy Consumption: Energy consumption means how much energy is consumed due to one-hop message exchange in

CDD and that of the flooding messages in SDD. According to the authors' observation the overall energetic cost of CDD is lower than that of SDD.

III. NODE REPLICATION ATTACK

Wireless sensor nodes are often deployed in hostile environments where they work unattended. Lacking tamper resistance due to low cost hardware components leave the nodes vulnerable to capture and compromise by an adversary. Thus a new type of attack called node replication attack arises in sensor networks. In this attack the adversary first analyze the captured node and then uses the credentials of the compromised node to introduce replicas at judiciously chosen network locations to launch a variety of insidious and hard-to-detect attacks.

Though replica nodes are controlled by the adversary, having legitimate information (codes, key materials) replicated from compromised nodes allow them to act like authorized participants of the network. If left undetected, node replication attack can easily subvert the main goal of the deployed sensor network by falsifying sensor data or suppressing legitimate data, extracting data from the network and staging denial of service attacks. Thus node replication attack is very dangerous and it is very important to develop software based counter measures to defend this attack.

A number of replica node detection schemes and protocols have been proposed for static sensor networks. However none of these schemes are suitable for mobile wireless sensor networks. A simple distributed solution for detecting node replication attacks in mobile wireless sensor networks can be designed by making some changes on LSM [6] or RED [7]. In these protocols location claims can be replaced by time-location claims which include the time when the claims are generated. Witnesses store all received time-location claims. After arriving a new time location claim to a witness, it is compared with all the old location claims to verify that whether the corresponding node is a replica. But this approach is not affordable by mobile sensor networks since the increased number of routing signal messages considerably reduce the lifetime of the network. In [8] Deng presented two mobility-assisted protocols (UTLSE and MTLSD) for detecting node replication attacks in mobile wireless sensor networks.

A. UTLSE: Unary-Time-Location Storage and Exchange

In this protocol each node in the sensor network is initialized with a unique tracking set, which means the node is a witness of each node in that tracking set. When a node meets with a new neighbor who is a member of its tracking set, it asks the neighbor to send a time-location claim to it. Meanwhile if the tracking set of the node and the neighbor is not disjoint and the ID of the node is bigger than its neighbor, it sends all the stored time-location claims of each common tracked node to its neighbor. If any witness receives two contradictory time-location claims for the same node identity (ID), it will have detected the existence of a replica and can take appropriate actions to revoke the node's credentials.

B. MTLSD: Multi-Time-Location Storage & Diffusion

To show that a loop hole exists in UTLSE protocol the following situation can be considered. Suppose two legitimate nodes a and b both are witnesses of a compromised node x . At time t_1 , node a encounters one replica of node x positioned at l_1 . At time t_2 , node b encounters another replica of node x positioned at l_2 . $\langle t_1, l_1 \rangle$ and $\langle t_2, l_2 \rangle$ are contradictory. However before node a encounters node b , they separately meet another replica of node x of which location is same (which is l_3). Then both of them replace l_1 and l_2 with l_3 . Thus node a (or node b) regards node x as a legitimate node. Though the described situation does not always occur, it reduces the detection probability to a certain extent. So MTLSD was introduced by making some changes to UTLSE to minimize the impact of loop hole. In MTLSD a FIFO queue of which length is three is maintained to store the corresponding location claims for each node in the tracking set. Assuming node a meets node b and their tracking set is not disjoint, both a and b send detection request messages to each other. Receiving these messages each of them insert the received time location claims different from that they have stored, into the corresponding queue at the right position. But like UTLSE only the node with smaller ID would launch the detection process.

C. Performance Analysis

Two metrics has been used by Deng to evaluate the efficiency of the protocols UTLSE and MTLSD.

Communication overhead: Communication overhead refers to the average number of the messages sent by a sensor node while propagating the location claims. According to the authors' calculation the communication overhead is $O(N)$ where N is the total number of nodes in the sensor network.

Storage overhead: Storage overhead is the average number of the location claims stored in a sensor node. Since each node tracks nodes, and for each tracked node, only one queue (with fixed length) is maintained for storing the corresponding location claims, the storage overhead of every node is O .

Detection probability and detection time (the delay between actual replica node deployment and detection) are two basic performance indices of UTLSE and MTLSD protocols. According to the authors' observation the detection probability of the MTLSD protocol is greater than the probability of protocol UTLSE. Besides simulation results show that when detection time is shorter, the MTLSD decreases its detection probability more than UTLSE do.

Because of the mobility-assisted property, the main advantage of these protocols is that they do not rely on any specific routing protocol, which makes them suitable for various mobile settings.

IV. WORMHOLE ATTACK

"For initiation a wormhole attack, an adversary connects two distant points in the network using a direct low-latency communication link called as the wormhole link. The wormhole link can be established by a variety of means, e.g., by using a Ethernet cable, a long-range wireless transmission, or an optical link. Once the wormhole link is established, the

adversary captures wireless transmissions on one end, sends them through the wormhole link and replays them at the other end" [9].

An example is shown in the figure 1. Here X and Y are the two end-points of the wormhole link (called as wormholes). X replays everything that Y hears in its neighborhood (area B) in its own neighborhood (area A) and vice versa. The net effect of such an attack is that all the nodes in area A assume that nodes in area B are their neighbors and vice versa. This, as a result, affects routing and other connectivity based protocols in the network. Once the new routes are established and the traffic in the network starts using the X - Y shortcut, the wormhole nodes can start dropping packets and cause network disruption. They can also spy on the packets going through and use the large amount of collected information to break any network security.

A. Mobile Sink Based Technique

The wormhole attacks or the collusion of malicious nodes can be minimized by using a mobile sink (MS) with multiple communication channels on a sensor network. To minimize the attack a new technique is proposed [10], which allows a MS to launch a secure link with any sensor node and protect against threats imposed by wormhole attacks and collusion of malicious nodes.

This technique [10] relies on the assumption that any physical device has only one radio which is incapable of simultaneously sending or receiving on more than one channel. At the time of network deployment, every sensor node is preloaded with polynomial shares of a randomly selected subset of polynomials, called the polynomial ring. The base station dispatches the MS to securely collect sensor data. The MS is loaded with its arbitrarily chosen subset of polynomials. Initially all the sensors and the MS have their radios tuned to a pre-selected common channel termed as discovery channel. Discovery channel is used by a sensor node to detect whether it is within propinquity of the MS.

The MS traverses the network that transmits beacon messages over the discovery channel; the beacon message contains the MS ID. Sensors that are in propinquity of the MS can perceive the MS beacons. The sensor node uses discovery channel 1 to establish both a common encryption key and a secure channel with the MS to transfer its encrypted data. The MS can establish a pair wise key with any sensor node on the fly.

For every sensor node u , which wants to communicate securely with the MS, the two must first establish a common key k between them. Second, the MS randomly picks a secure channel f_i from a set of c channels $\{f_1, f_2, f_3 \dots f_{c-1}, f_c\}$. The MS then sends the message $\{f_i\}_k$ over the discovery channel to sensor node u . The sensor node used the shared key k to decrypt the encrypted message and the node u uses this secret channel for a specified period of T_s seconds. Node u transmits the encrypted data message $\{data\}_k$ to the MS by using the secret channel f_i .

The sensor node u switches back from radio to the discovery channel after T_s seconds. The MS picks a channel f_j from the randomly assigned list and turns its radio from the

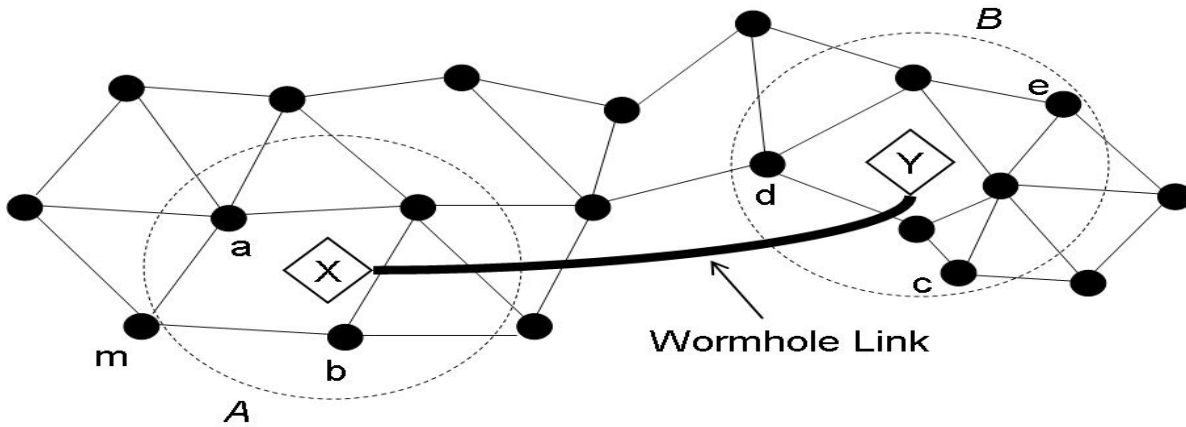


Figure 1: Wormhole attack.

discovery channel to f_j for a specified period of t_s sec, where $t_s \ll T_s$. The MS transmits beacon messages over the channel f_j and listens to it. Sensor nodes that are in the propinquity of the MS's range and have their radios tuned in to f_j hears the MS transmission and reply back by sending their encrypted data messages to the MS. If MS didn't find any sensor node tuning to the advertise channel then MS deletes this channel from the list of assigned channels. After time t_s , the MS tunes its radio back to the discovery channel and transmits beacons that contain its ID, so that sensor nodes that were not in the MS's range before and now are within range will be able to establish a secure communication link with the MS. Similarly, the process is repeated with every channel chosen by the MS from the list of assigned channels.

B. Performance Analysis

With this technology [10] a security scheme for WSN with MS is presented and observed that even when 50% of a sensor node's neighbor is malicious the single extra channel for communication with the MS brings down the probability of wormhole attack down to zero and improve the resilience of the network against wormhole attacks and node collusion.

V. SINKHOLE ATTACK

"It is a type of attack which has more than one malicious node of attackers make a compromised node looks more attractive to surrounding nodes by forging routing information"[11]. A sinkhole attack prevents the base station from taking complete and correct data, which may cause severe threat to higher-layer applications [12]. In a Sinkhole attack [12], a compromised node tries to illustrate all or as much traffic as possible from a particular region. It makes itself look eye-catching to the adjacent nodes with respect to the routing metric. Consequently, the adversary manages to attract all traffic that is intended to the base station. It take part in the routing process then it can launch more severe attacks, like selective forwarding, modifying or even dropping the packets coming through around.

A. Application of Secure Path Redundancy Protocol

Only few routing protocol is present such as SPINS and PRSA for WSN to protect against different attacks. It is found that homogenous network suffer from poor fundamental limit and performance. To get the better performance, heterogeneous model is more suitable. That's why PRSA model is more suitable for HSN by incorporating alternative path and mobility model for mobile sink to defend the node from sinkhole attack in HSN. In HSN few high sensor (H-Sensor) and large number of low sensor (L-sensor) is present.

The alternative path algorithm [11] can applied to various networks consists of different number of nodes, network capacity and different attacks. It is assumed that nodes are mobile in the network. The main objective of the protocol is finding the secure multiple paths between source and destination in the occurrence of sinkhole attacks.

B. Mechanism of SPR

In case of sinkhole attack the surrounding node choose the compromised node to pass the routing information. This is done by removing one or more active adversary node from the routing path. This algorithm [11] is used to detect such types of nodes by using a set of parameters (e.g. packet id, no of hop count, delay). It actually reflects the presence of adversary nodes. This secured mechanism can defend against attacks such as sinkhole attacks.

It is found that using mobile sink each of its location continuously disseminate throughout the entire sensor network to inform each of the sensor node about the direction of forwarding future data. Unfortunately frequent update of location from multiple sink leads to excessive communication of battery supply and increased collision in wireless transmission. But by using the secured path redundancy algorithm for HSN approach can avoid these limitations. The reason behind is that it consume less energy, better delay and efficient data delivery to more than one mobile sink.

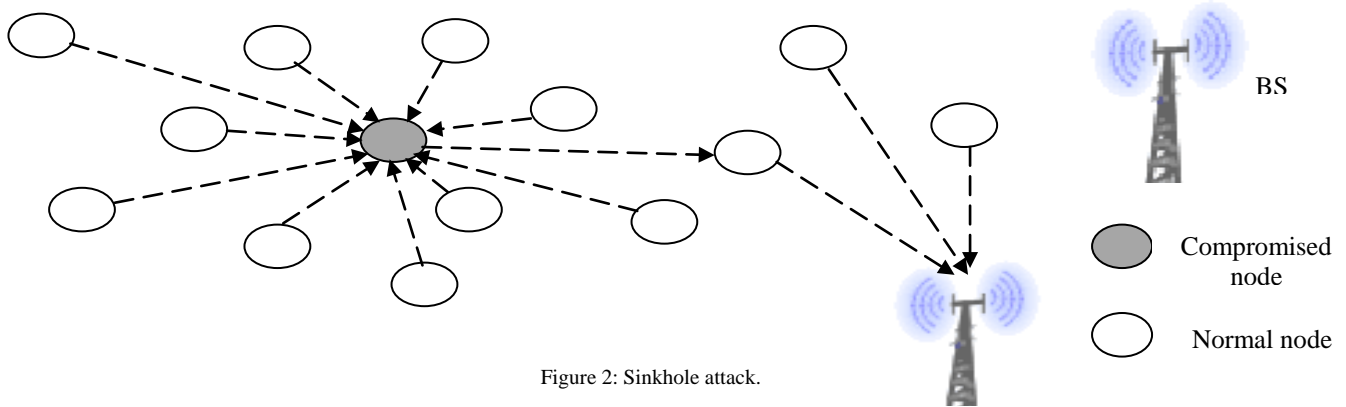


Figure 2: Sinkhole attack.

VI. CONCLUSION

Wireless sensor networks are emerging technology with many important applications. It is envisioned that sensor networks will be used in critical infrastructure in future. So security is essential to the success of applying WSN. Past researches on security has been focused on static sensor networks. The inherent mobility feature of wireless sensor network make it vulnerable to threats, and that solutions developed for static sensor networks are often either unsuitable or not directly applicable to the mobile wireless sensor networks. In this paper typical attacks like node compromise attack, node replica attack, sinkhole and wormhole attacks on mobile sensor networks have been summarized and the literatures on several countermeasures have been surveyed. Many security issues relevant to mobile WSNs remain open and more research activities on these exciting topics are expected to be covered in the future.

REFERENCES

- [1] Alexander Becher; Er Becher; Zinaida Benenson, Maximillian Dornseif; "Tampering with Motes: Real-World Physical Attacks on Wireless Sensor Networks"; Proceeding of the 3rd International Conference on Security in Pervasive Computing (SPC), 2006; pages 104–118.
 - [2] J. Newsome, E. Shi, D. Song, and A. Perrig. The Sybil attack in sensor networks: analysis & defenses. In IPSN'04, pages 259–268, 2004.
 - [3] B. Parno, A. Perrig, and V. Gligor, "Distributed detection of node replication attacks in sensor networks," in Proc. 2005 IEEE Symposium on Security and Privacy, Oakland, CA, USA, May 2005, pp. 49–63.
 - [4] Wei Ding; Laha, B.; Yenduri, S., "First stage detection of compromised nodes in sensor networks," *Sensors Applications Symposium (SAS)*, 2010 IEEE, vol., no., pp.20-24, 23-25 Feb. 2010.
 - [5] Mauro Conti, Roberto Di Pietro, Luigi V. Mancini, and Alessandro Mei. Emergent Properties: Detection of the Node-capture Attack in Mobile Wireless Sensor Networks. In Proceedings of the First ACM Conference on Wireless Network Security (ACM SIGSAC WiSec 2008), pages 214–219, Alexandria, VA, USA, March 31 - April 2, 2008.
 - [6] Parno, B., Perrig, A., and Gligor, V. 2005. Distributed detection of node replication attacks in sensor networks, Security and Privacy, 2005 IEEE Symposium on, vol., no., and pp. 49-63.
 - [7] Conti, M., Di Pietro, R., Mancini, L. V., and Mei, A. 2007. A randomized, efficient, and distributed protocol for the detection of node replication attacks in wireless sensor networks. In Proceedings of the 8th ACM international Symposium on Mobile Ad Hoc Networking and Computing (Montreal, Quebec, Canada, September 09 - 14, 2007). MobiHoc '07. ACM, New York, NY, 80-89.
 - [8] Xiaoming Deng; Yan Xiong; Depin Chen; "MoBility-Assisted Detection of the Replication attacks in mobile wireless sensor networks," *Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2010 IEEE 6th International Conference on, vol., no., pp.225-232, 11-13 Oct. 2010.
 - [9] Detecting Wormhole Attacks
<http://www.wings.cs.sunysb.edu/~ritesh/wormhole.html>. [Accessed Feb 2011].
 - [10] Rasheed, A.; Mahapatra, R.; "Mobile Sink Using Multiple Channels to Defend Against Wormhole Attacks in Wireless Sensor Networks" (IPCCC), 2009 IEEE 28th International, pp- 216 – 222.
 - [11] P. Samundiswary, Padma Priyadarshini and P. Dananjayan, "Detection of Sinkhole Attacks for Mobile Nodes in Heterogeneous Sensor Networks with Mobile Sinks" International Journal of Computer and Electrical Engineering, Vol. 2, No. 1, February, 2010
 - [12] Krontiris, I.; Giannetos, T.; Dimitriou, T., "Launching a Sinkhole Attack in Wireless Sensor Networks; The Intruder Side", IEEE International Conference on Wireless & Mobile Computing, Networking & Communication 2008 wimob, pp.526-531
- Cheng-Lung Yang, Wernhuar Tarng, Kuen-Rong Hsieh and Mingteh Chen, "A Security Mechanism for Clustered Wireless Sensor Networks Based on Elliptic Curve Cryptography" IEEE e news letter, intelligent internet system issue #33 December 2010.

AUTHORS PROFILE

Abu Saleh Md. Tayeen has been serving as a Lecturer at the Department of Computer Science of American International University-Bangladesh (AIUB), Dhaka, Bangladesh since September 2008. Before that he worked as a Software Engineer in Ice Technologies & Services Ltd. from April 2007 to March 2008.

A.F.M Sultanul Kabir is currently working as a Assistant Professor in American International University of Bangladesh (AIUB), Dhaka Bangladesh. During the period June 2008-May 2009 he worked as a researcher in Swedish Defence Research agency (FOI), Stockholm Sweden.

Razib Hayat Khan is doing his PhD at Department of Telematics, Norwegian University of Science and Technology (NTNU), Norway. He worked as visiting researcher at Duke University, Durham, USA and served as research engineer, Multimedia technologies at Ericsson AB, Sweden. At present he is working with performance and dependability issues in Communication system.

Hybrid Multi-level Intrusion Detection System

Sahar Selim, Mohamed Hashem and Taymoor M. Nazmy

Faculty of Computer and Information Science

Ain Shams University

Cairo, Egypt

Sahar.Soussa@gmail.com

Abstract— Intrusion detection is a critical process in network security. Nowadays new intelligent techniques have been used to improve the intrusion detection process. This paper proposes a hybrid intelligent intrusion detection system to improve the detection rate for known and unknown attacks. We examined different neural network & decision tree techniques. The proposed model consists of multi-level based on hybrid neural network and decision tree. Each level is implemented with the technique which gave best experimental results. From our experimental results with different network data, our model achieves correct classification rate of 93.2%, average detection rate about 95.6%; 99.5% for known attacks and 87% for new unknown attacks, and 9.4% false alarm rate.

Keywords—component; network intrusion detection; neural network; Decision Tree; NSL-KDD dataset

I. INTRODUCTION

Security of network system is becoming increasingly important as more sensitive information is being stored and manipulated online. It is difficult to prevent attacks only by passive security policies, firewall, or other mechanisms. Intrusion Detection Systems (IDS) have thus become a critical technology to help protect these systems as an active way. An IDS can collect system and network activity data, and analyze those gathered information to determine whether there is an attack [1].

The main objective of this work is to design and develop security architecture (an intrusion detection and prevention system) for computer networks. This proposed system should be positioned at the network server to monitor all passing data packets and determine suspicious connections. Therefore, it can inform the system administrator with the suspicious attack type. Moreover, the proposed system is adaptive by allowing new attack types to be defined.

We build the model to improve the detection rate for known and unknown attacks. First, we train and test our hybrid model on the normal and the known intrusion data. Then we test our system for unknown attacks by introducing new types of attacks that are never seen by the training module.

II. PREVIOUS WORK

An increasing amount of research has been conducted for detecting network intrusions. The idea behind the application

of soft computing techniques in implementing IDSs is to include an intelligent agent in the system that is capable of disclosing the latent patterns in abnormal and normal connection audit records, and to generalize the patterns to new (and slightly different) connection records of the same class.

There are researches that implement an IDS using Multi-layer perceptron (MLP) which have the capability of detecting normal and attacks connection as in [2], [3]. Reference [4] used MLP not only for detecting normal and attacks connection but also identify attack type.

Decision Tree (C4.5 Algorithm) was explored as intrusion detection models in [5] and [6].

Neural network and C4.5 have different classification capabilities for different intrusions. Therefore, Hybrid model improves the performance to detect intrusions. [1], [7] compare the performance of Hybrid model, single Back Propagation network, and single C4.5 algorithm. Experimental results demonstrate that neural networks are very interesting for generalization and very poor for new attacks while decision trees have proven their efficiency in both generalization and new attacks detection. A multi-classifier model, where a specific detection algorithm is associated with an attack category for which it is the most promising, was built in [8].

Reference [9] developed a multi-stage neural network which consists of three detection levels. The first level differentiates between normal and attack. The second level specifies whether this attack is DOS or probe. The third detection level identifies attacks of denial of service and probe attacks.

The proposed system is a hybrid multi-level system. It consists of three levels. Each level was examined with different machine learning techniques. Each module in each level is built using the best classifier which gave best results for this level. It has the ability to identify normal and attack records and also being able to detect attack type by the next levels. This approach has the advantage to flag for suspicious record even if attack type of this record wasn't identified correctly.

III. THE PROPOSED SYSTEM

Our system is a modular network-based intrusion detection system that analyzes Tcpdump data using data mining techniques to classify the network records to not only normal and attack but also identify attack type.

The main characteristics of our system:

- **Multilevel:** has the capability of classifying network intruders into a set of different levels. The first level classifies the network records to either normal or attack. The second level can identify four categories/classes. The third level where the attack type of each class can be identified.

Attacks of the same class have a defined signature which differentiates between attacks of every class/category from others, i.e. DOS attacks have similar characteristics which identifies them from attacks of Probing, R2L and U2R. That's why there's often misclassification between attacks of the same class, which gave the importance of making a multi-stage system consisting of three levels.

The data is input in the first level which identifies if this record is a normal record or attack. If the record is identified as an attack then the module would raise a flag to the administrator that the coming record is an attack then the module inputs this record to the second level which identifies the class of the coming attack. Level 2 module pass each attack record according to its class type to level 3 modules. Level 3 consists of 4 modules one for

each class type (DOS, Probe, R2L, U2R). Each module is responsible for identifying the attack type of coming record.

The idea is that if ever the attack name of the third level is misclassified then at least the admin was identified that this record is suspicious after the first level network. Finally the admin would be alerted of the suspected attack type to guide him for the suitable attack response [9].

- **Hybrid:** Modules of each level can use different data mining technique. We made a comparative study examining several data mining techniques to find the best classifier for each level. Neural network and decision trees have different classifying abilities for different intrusions. Neural network have high performance to DOS and Probing attacks while decision trees can detect the R2L more accurately than neural network. Therefore, Hybrid model will improve the performance to detect intrusions.

- **Adaptive:** Attacks that are misclassified by the IDS as normal activities or given wrong attack type will be relabeled by the network administrator. The training module can be retrained at any point of time which makes its implementation adaptive to any new environment and/or any new attacks in the network.

IV. SYSTEM ARCHITECTURE

The system components as shown in Fig 1 are:

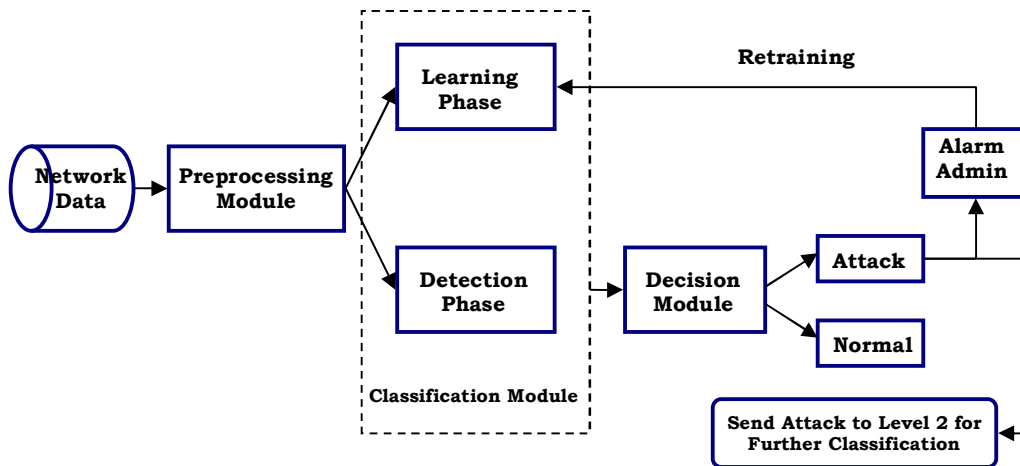


Figure 1. System architecture

A. The Capture Module

Raw data of the network are captured and stored using the network adapter.

B. The Preprocessing Module

This module is responsible for Numerical Representation, Normalization and Features selection of raw input data to be used by the classification module. The preprocessing module

maps the raw packets captured from the network by the TCP dump capture utility to a set of patterns of the most Effective Selected Feature. These dominant features are then used as inputs to the training module.

The preprocessing module consists of three phases: [9]

- 1) **Numerical Representation:** Converts non-numeric features into a standardized numeric representation. This process involved the creation of relational tables for each of

the data type and assigning number to each unique type of element. (e.g. protocol_type feature is encoded according to IP protocol field: TCP=0, UDP=1, ICMP=2). This is achieved by creating a transformation table containing each text/string feature and its corresponding numeric value.

2) *Normalization*: The ranges of the features were different and this made them incomparable. Some of the features had binary values where some others had a continuous numerical range (such as duration of connection). As a result, inputs to the classification module should be scaled to fall between zero and one [0, 1] range for each feature.

3) *Dimension reduction*: reduce the dimensionality of input features of the classification module. Reducing the input dimensionality will reduce the complexity of the classification module, and hence the training time.

C. The classification Module

The classification module has two phases of operation. The learning and the detection phase.

1) The Learning Phase

In the learning phase, the classifier uses the pre-processed captured network user profiles as input training patterns. This phase continues until a satisfactory correct classification rate is obtained.

2) The Detection Phase

Once the classifier is learned, its capability of generalization to correctly identify the different types of users should be utilized to detect intruder. This detection process can be viewed as a classification of input patterns to either normal or attack.

D. The Decision Module

The basic responsibility of the decision module is to transmit alert to the system administrator informing him of coming attack. This gives the system administrator the ability to monitor the progress of the detection module.

1) Performance Measures

To evaluate our system we used two major indices of performance. We calculate the detection rate and the false alarm rate according to [10] the following assumptions:

- False Positive (FP): the total number of normal records that are classified as anomalous
- False Negative (FN): the total number of anomalous records that are classified as normal
- Total Normal (TN): the total number of normal records
- Total Attack (TA): the total number of attack records
- Detection Rate = $[(TA-FN) / TA] * 100$
- False Alarm Rate = $[FP/TN] * 100$
- Correct Classification Rate = Number of Records Correctly Classified / Total Number of records in the used dataset

V. MACHINE LEARNING ALGORITHMS APPLIED TO INTRUSION DETECTION

Seven distinct pattern recognition and machine learning algorithms were tested on the NSL-KDD dataset. These algorithms were selected in the fields of neural networks and decision trees.

A. Neural Networks

The neural network gains the experience initially by training the system to correctly identify pre-selected examples of the problem. The response of the neural network is reviewed and the configuration of the system is refined until the neural network's analysis of the training data reaches a satisfactory level. In addition to the initial training period, the neural network also gains experience over time as it conducts analysis on data related to the problem [2].

1) Multi-Layer Perceptron (MLP)

The architecture used for the MLP during simulations consisted of a three layer feed-forward neural network: one input, two hidden, and one output layers. Sigmoid transfer functions were used for each neuron in both the hidden layers and softmax in the output layers. The network was set to train until the desired mean square error of 0.001 was met or 10000 epochs was reached.

For the first level there were 31 neurons in the input layer (31-feature input pattern) after feature selection, 22 neurons in first hidden layer, 18 neurons in second hidden layer and 2 neurons (one for normal and the other for attack) in the output layer. During the training process, the mean square error is 0.0157 at 10000 epochs. For the second level 38 in input layer, 12 in first hidden layer, 10 in second hidden layer and 4 neurons in the output layer (DOS, Probe, R2L and U2R). During the training process, the mean square error is 0.0114 at 10000 epochs. We've four networks in the third level. DOS network has layers of 28-2-2-7 feed-forward neural network. (i.e. 28 in input layer, 2 in the 1st hidden layer, 2 in the 2nd hidden layer and 7 in the output layer). During the training process, the mean square error is 0 at 1574 epochs. Probe network has layers of 24-22-14-6 feed-forward network with mean square error 0.05 at 10000 epochs. R2L network has layers of 26-17-10-5 feed-forward network with mean square error 0 at 5838 epochs. U2R network has layers of 11-9-7-5 feed-forward network with mean square error 2.33 at 10000 epochs.

2) Radial Basis Function (RBF)

The RBF layer uses Gaussian transfer functions. The learning rate was set to 0.1 for the hidden layer and 0.01 for the output layer. The alpha was set to 0.75. For the first level there were 31 neurons in the input layer, 10 neurons in hidden layer and 2 neurons (one for normal and the other for attack) in the output layer. Estimated accuracy of training was 94.4%. The second level has 37 in input layer, 10 in hidden layer and 4 neurons in the output layer (DOS, Probe, R2L and U2R) with estimated accuracy of 93.5%. We've four networks in the third level. DOS RBF network has layers of 28-20-7. (i.e. 28 in input layer, 20 in hidden layer and 7 in the output layer) with estimated accuracy 100%. Probe network has layers of 24-20-6 network with estimated

accuracy 98.3%. R2L RBF network has layers of 26-20-5 with estimated accuracy 98.3%. U2R network has layers of 11-20-5 with estimated accuracy 75%.

3) Exhaustive Prune

The first level there consists of 13 neurons in the input layer, 22 neurons in first hidden layer, 7 neurons in second hidden layer and 2 neurons (one for normal and the other for attack) in the output layer with estimated accuracy of training 99.8%. The second level consists of 25 in input layer, 9 in first hidden layer, 5 in second hidden layer and 4 neurons in the output layer (DOS, Probe, R2L and U2R) with accuracy of training 99.9%. We've four networks in the third level. DOS network has layers of 3-19-17-7 network with accuracy of training 100%. Probe network has layers of 10-12-5-6 network with estimated accuracy of 99.6%. R2L network has layers of 14-3-2-5 network with estimated accuracy of 100%. U2R network has layers of 1-3-2-5 network with estimated accuracy of training 81.5%.

B. Decision trees

The decision tree is a simple if then else rules but it is a very powerful classifier and proved to have a high detection rate. They are used to classify data with common attributes. Each decision tree represents a rule which categorizes data according to these attributes. A decision tree consists of nodes, leaves, and edges. A node of a decision tree specifies an attribute by which the data is to be partitioned. Each node has a number of edges which are labeled according to a possible value of the attribute in the parent node. An edge connects either two nodes or a node and a leaf. Leaves are labeled with a decision value for categorization of the data [11].

1) C5

See5.0 (C5.0) is one of the most popular inductive learning tools originally proposed by J.R.Quinlan as C4.5 algorithm (Quinlan, 1993) [11]. Single C5 acquires pruned decision tree with pruning severity 75% and winnowing attributes. First level consists of 121 nodes on train data and 20 tree depth and standard error 0.01%. Second level consists of 113 nodes and tree depth of 12 with standard error 0.05%. Third level DOS tree consists of 6 nodes and tree depth of 4 levels with standard error 0%. Probe tree consists of 69 nodes and tree depth of 10 levels with standard error 0.4%. R2L tree consists of 7 nodes and tree depth of 4 levels with standard error 0%. U2R tree consists of 9 nodes and tree depth of 4 levels with standard error 8.33%.

2) Classification and Regression Trees (CRT or CART)

CRT was set of maximum surrogates 10, minimum change in impurity 0.0 and Gini impurity measure for categorical targets. First level consists of 15 nodes and of depth 4. Second level consists of 15 nodes of tree depth 4. Third level DOS consists of 7 nodes of tree depth = 3. Probe consists of 13 nodes of tree depth 5. R2L consists of 7 nodes of tree depth 4. U2R consists of 17 nodes of tree depth 6.

3) Chi-squared Automatic Interaction Detector (CHAID)

CHAID was adjusted of Alpha splitting 0.05, alpha for merging 0.05, epsilon for convergence 0.001, using pearson

chi-square method. First level consists of 35 nodes and of depth 5. Second level consists of 28 nodes of tree depth 4. Third level DOS consists of 6 nodes of tree depth 3. Probe consists of 49 nodes of tree depth 6. R2L consists of 7 nodes of tree depth 3. U2R consists of 12 nodes of tree depth 5.

4) Quick, Unbiased, Efficient Statistical Tree (QUEST)

QUEST was adjusted of maximum surrogates 5, and alpha for splitting 0.05. First Level consists of 15 nodes and of 4 tree depth. Third level DOS consists of 11 nodes of tree depth 6. Probe consists of 17 nodes of tree depth 6. R2L consists of 9 nodes of tree depth 5. U2R consists of 13 nodes of tree depth 6.

VI. EXPERIMENTS AND RESULTS

A. Dataset Description

KDDCUP'99 is the mostly widely used data set for the evaluation of these systems. The KDD Cup 1999 uses a version of the data on which the 1998 DARPA Intrusion Detection Evaluation Program was performed. They set up an environment to acquire raw TCP/IP dump data for a local-area network (LAN) simulating a typical U.S.Air Force LAN.

1) *There are four major categories of networking attacks. Every attack on a network can be placed into one of these groupings [13].*

a) *Denial of Service Attack (DoS):* is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. e.g. apache, smurf, Neptune, ping of death, back, mail bomb, UDP storm, etc.

b) *User to Root Attack (U2R):* is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system. e.g. xlock, guest, xnsnoop, phf, sendmail dictionary etc.

c) *Remote to Local Attack (R2L):* occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine. e.g. perl, xterm.

d) *Probing Attack:* is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls. e.g. satan, saint, portsweep, mscan, nmap etc.

There are some inherent problems in the KDDCUP'99 data set [12], which is widely used as one of the few publicly available data sets for network-based anomaly detection systems. The first important deficiency in the KDD data set is the huge number of redundant records. Analyzing KDD train and test sets, it was found that about 78% and 75% of the records are duplicated in the train and test set, respectively. This large amount of redundant records in the

train set will cause learning algorithms to be biased towards the more frequent records, and thus prevent it from learning infrequent records which are usually more harmful to networks such as U2R attacks. The existence of these repeated records in the test set, on the other hand, will cause the evaluation results to be biased by the methods which have better detection rates on the frequent records [13].

The data in the experiment is acquired from the NSL-KDD dataset which consists of selected records of the complete KDD data set and does not suffer from mentioned shortcomings by removing all the repeated records in the entire KDD train and test set, and kept only one copy of each record [13]. Although, the proposed data set still suffers from some of the problems and may not be a perfect representative of existing real networks, because of the lack of public data sets for network-based IDSs, but still it can be

applied as an effective benchmark data set to help researchers compare different intrusion detection methods. The NSL-KDD dataset is available at [14].

In this study we examine using attacks from the four classes to check the ability of the intrusion detection system to identify attacks from different categories. The sample dataset contains 83655 record for training (40000 normal and 43655 for attacks) and 16592 for testing (9657 normal, 6935 for known attacks and 3202 for unknown attacks).

B. Level 1 output

Level 1 duty is to classify whether coming record is normal or attack. It is observed that MLP best classifies normal records while C5 is more efficient in detecting known and unknown attacks. The results of Level 1 are shown in table 1 & 2.

TABLE I. CORRECT CLASSIFICATION RATE FOR LEVEL 1

Percentage	Normal	Attacks	New Attacks	Correct Classification Rate
MLP	95.1	97.2	78.7	93.2
RBF	90.4	93.1	45.5	84.1
Exhaustive	89.7	97.3	86.2	91.8
C5	90.6	99.5	97	93.2
CRT	93.3	98.9	45.4	87.5
QUEST	85.5	98	67.1	86.9
CHAID	89.6	97.1	59.2	87.3

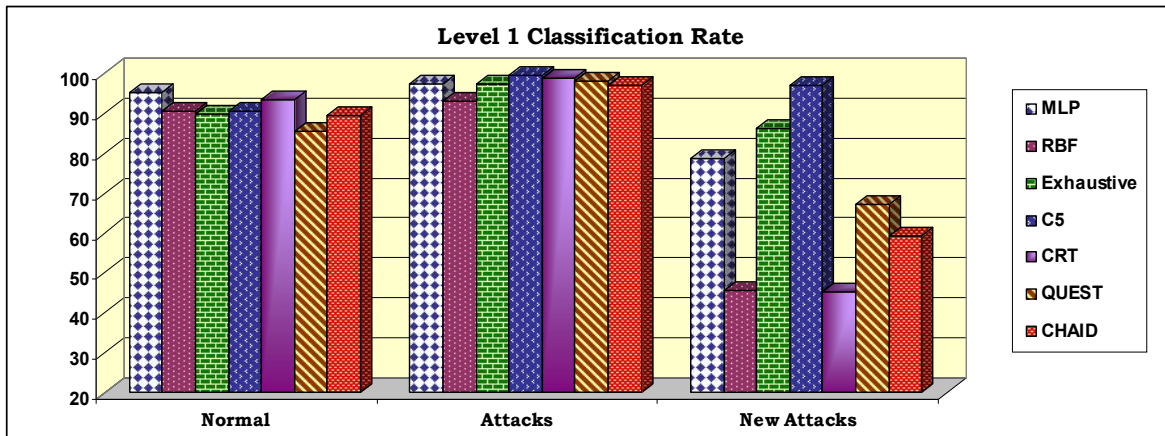


Figure 2. Level 1 Classification Rate

TABLE II. DETECTION RATE & FALSE ALARM RATE FOR LEVEL 1

Classifier	Detection Rate	False Alarm Rate
MLP	91.397	5
RBF	78.0979	9.64
Exhaustive	91.83	10.3
C5	95.5702	9.4
CRT	82.0343	15.8
QUEST	88.2301	14.53
CHAID	85.1322	10.44

C5 has a significant detection rate for known and unknown attacks but it produce higher false alarm rate compared to MLP.

C. Level 2 Output

Records classified as attacks by the first level are introduced to second level which is responsible for classifying coming attack to one of the four classes (DOS, Probe, R2L & U2R). Testing results showed that C5 & CRT (decision trees) produced best correct classification rate for second level as shown in table 3.

TABLE III. CORRECT CLASSIFICATION RATE FOR LEVEL 2

Level 2 Classifiers	Known Attacks	New Attacks	Correct Classification
MLP	82.8202	56.2637	82.8202
RBF	74.7977	50.6717	74.7977
Exhaustive	79.2382	49.8594	79.2382
C5	86.0174	59.294	86.0174
CRT	85.7805	62.6679	85.7805
CHAID	78.7646	38.8316	78.7646

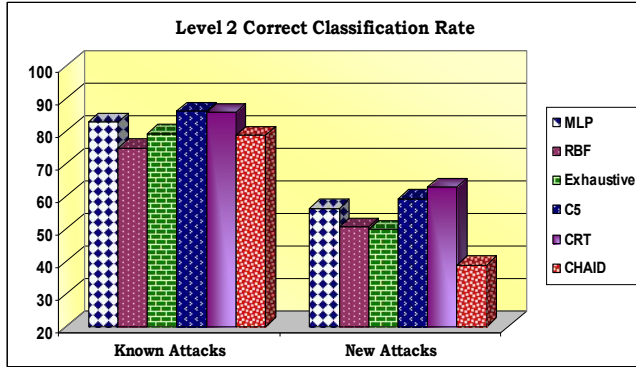


Figure 3. Level 2 Classification Rate

D. Level 3 Output

The third level consists of four modules; a module for each class. For example records that were classified by the second level to be DOS attack are sent to the DOS module of the 3rd level & so on.

Results of Denial of service modules showed that DOS attacks are easy to be correctly classified by many classifiers either neural network or decision trees as shown in table 4.

TABLE IV. DOS ATTACKS CLASSIFICATION RATE

DOS Classifier	Correct Classification Rate
MLP	100
RBF	99.3852
Exhaustive	99.9297
C5	100
CRT	100
QUEST	99.9297
CHAID	100

Results of Probe module showed that C5 & MLP are most efficient for detecting this type of attacks as shown in table 5.

TABLE V. PROBE ATTACKS CLASSIFICATION RATE

Probe Classifier	Correct Classification Rate
MLP	99.3
RBF	97.8
Exhaustive	97
C5	98.6
CRT	92.6

QUEST	94.1
CHAID	95.5

Results of R2L module showed that C5 are most efficient for detecting this type of attacks significantly as shown in table 6.

TABLE VI. R2L ATTACKS CLASSIFICATION RATE

R2L Classifier	Correct Classification Rate
MLP	91
RBF	93
Exhaustive	91
C5	100
CRT	97
QUEST	96
CHAID	97

U2R attacks have a very low classification rate compared to other classes. Results showed that Exhaustive prune is better than other classifiers for detecting attacks of this class as shown in table 7.

TABLE VII. U2R ATTACKS CLASSIFICATION RATE

U2R Classifier	Correct Classification Rate
MLP	48.2
RBF	43.1
Exhaustive	54.4
C5	44.1
CRT	44.1
QUEST	35.3
CHAID	41.2

VII. DISCUSSION

Simulation results demonstrated that for a given attack category certain classifier algorithms performed better. Consequently, a multi-classifier model that was built using most promising classifiers for a given attack category was evaluated for probing, denial-of-service, user-to-root, and remote-to-local attack categories.

While the neural networks are very interesting for generalization and very poor for new attacks detection, the decision trees have proven their efficiency in both generalization and new attacks detection. Besides the C5 has less training time than the MLP. However, none of the machine learning classifier algorithms evaluated was able to perform detection of user-to-root attack categories significantly (no more than 54% detection for U2R category).

The advantage of the proposed multi-level system is not only higher accuracy but also the parallelism as every module can be trained on separate computer which provides less training time. Also the multi-level powers the system with scalability because if new attacks of specific class are added to the dataset we don't have to train all the modules but only the module affected by the new attack. Attacks that are misclassified by the IDS as normal activities or given

wrong attack type will be relabeled by the network administrator. Training module can be retrained at any point of time which makes its implementation adaptive to any new environment or any new attacks in the network.

VIII. CONCLUSION & FUTURE WORK

In this paper we develop a hybrid multilevel intrusion detection system. The proposed system consists of three detection levels. The network data are introduced to the module of the first level which aims to differentiate between normal and attack. If the input record was identified as an attack then the administrator would be alarmed that the coming record is suspicious and then this suspicious record would be introduced to the second level which specifies the class of this attack (DOS, probe, R2L or U2R). The third detection level consists of four modules one module for each class type to identify attacks of this class. Finally the administrator would be alarmed of the expected attack type [9].

We examined each module using different machine learning models (MLP, RBF, C5, CRT, QUEST & Exhaustive Prune). Each module is implemented with the most promising classifier that gave highest correct classification rate. Therefore, Hybrid model will improve the performance of intrusion detection.

The experimental results show that the designed multi-level system has detection rate equal to 95.6% for both (known and unknown attacks). The first level is implemented by C5 decision tree which showed significant detection rate for both known and unknown attacks. The drawback of using C5 decision tree is the high false alarm rate that it produces. The second level is implemented by C5. As for the third level DOS & Probe modules are implemented by MLP, R2L module is implemented by C5 decision tree and U2R module is implemented by Exhaustive prune.

The detection of U2R attack is more difficult because of their close resemblance with the normal connections. Our future research will be directed towards developing more accurate base classifiers particularly for the detection of U2R attacks. Also finding ways to produce less false alarm rate for the C5 Decision tree.

REFERENCES

- [1] Z.S. Pan, S.C. Chen, G.B Hu and D.Q. Zhang, "Hybrid Neural Network and C4.5 for Misuse Detection," In Machine Learning and Cybernetics, pp. 2463-2467. Xi'an, 2003.
- [2] J.Cannady, "Artificial neural networks for misuse detection," Proceedings of the 1998 National Information Systems Security Conference (NISSC'98), Arlington, VA, pp. 443-456, 1998.
- [3] Srinivas Mukkamala, "Intrusion detection using neural networks and support vector machine," Proceedings of the 2002 IEEE International Honolulu, HI, 2002.
- [4] M. Moradi, and M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks," IEEE International Conference on Advances in Intelligent Systems - Theory

and Applications, Luxembourg-Kirchberg, Luxembourg, November 15-18, 2004.

- [5] Dewan Md. Farid, Nouria Harbi, Emna Bahri, Mohammad Zahidur Rahman and Chowdhury Mofizur Rahman, "Attacks Classification in Adaptive Intrusion Detection using Decision Tree," International Conference on Computer Science (ICCS 2010), 29-31 March, 2010, Rio De Janeiro, Brazil.
- [6] L Prema RAJESWARI and Kannan ARPUTHARAJ, "An Active Rule Approach for Network Intrusion Detection with Enhanced C4.5 Algorithm," International Journal of Communications, Network and Systems Sciences (IJCNSS), 2008, 4, 285-385.
- [7] Y. Bouzida, F.Cuppens, "Neural networks vs. decision trees for intrusion detection," IEEE/IST Workshop on Monitoring, Attack Detection and Mitigation (MonAM), Tuebingen, Germany, 28-29 September 2006.
- [8] M.R. Sabhnani and G. Serpen, "Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context," Proceedings of International Conference on Machine Learning: Models, Technologies, and Applications, Las Vegas, Nevada, 2003, pp. 209-215.
- [9] Sahar Selim, M. Hashem and Taymoor M. Nazmy, "Intrusion Detection using Multi-Stage Neural Network," International Journal of Computer Science and Information Security, Vol. 8, No. 4, 2010.
- [10] S.T. Sarasamma, Q.A. Zhu, and J. Huff, "Hierarchical Kohonen Net for Anomaly Detection in Network Security," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 35(2), 2005, pp. 302-312.
- [11] Quinlan JR. "C4.5: programs for machine learning," Log Altos, CA: Morgan Kaufmann; 1993.
- [12] KDD Cup 1999. Available on: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>, October 2007
- [13] M. Tavallaei, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.
- [14] "NSL-KDD data set for network-based intrusion detection systems," Available on: <http://nsl.cs.unb.ca/NSL-KDD/>, March 2009

AUTHORS PROFILE

Sahar Selim Fouad Bachelor of Computer Science, Faculty of Computer & Information Science, Ain Shams University. Currently working for master degree. Fields of interest are intrusion detection, computer and networks security.

Mohamed Abdel-Aziz Hashem Professor in IT and Security, Ain Shams University. Currently Vice Dean of Educational & Students' Affairs, faculty of Computer and Information Science, Ain Shams University. Fields of interest are computer networks, Ad-hoc and wireless networks, Qos Routing of wired and wireless networks, Modeling and simulation of computer networks, VANET and computer and network security.

Taymoor Mohammed Nazmy Professor in Computer Science, Ain Shams University. He served before in faculties of Sciences, and education as a lecturer for over 12 years. He was the director of the university information network. Currently Vice Dean of higher studies and researches, faculty of Computer and Information Science, since 2007. Fields of interest are image processing, pattern recognition, artificial neural networks, networks security and speech signal analysis.

Skin Lesion Segmentation Algorithms using Edge Detectors

J.H.Jaseema Yasmin¹

Associate Professor, Department of Computer Science
and Engineering
National College of Engineering, Tirunelveli, India.
jaseemay@yahoo.co.in

M.Mohamed Sathik²

Associate Professor in Computer Science
Sadakathullah Appa College, Tirunelveli – India
mmdsadiq@gmail.com

Abstract— An effective segmentation algorithm using log edge detector, for border detection of real skin lesions is presented which insinuate the excessive growth or regression of a melanoma, that helps in early detection of malignant melanoma and its performance is compared with the segmentation algorithm using canny detector, developed by us previously for border detection of real skin lesions. The experimental results demonstrate the successful border detection of noisy real skin lesions by the effective segmentation algorithm using log detector. We conclude that the segmentation algorithm using log detector, segments the lesion from the image even in the presence of noise for a variety of lesions, and skin types and its performance is better than the segmentation algorithm that we have developed previously that uses canny detector, for border detection of real skin lesions for noisy skin lesion diagnosis.

Keywords- Segmentation; Skin Lesion; log edge detector; canny edge detector; Border detection; Melanoma.

I. INTRODUCTION

Image segmentation is used to locate objects and boundaries in images, is not a simple task due to the great variety of lesions, skin types, presence of hair etc [14].

Once a image is selected , the system should provide an automatic identification (or segmentation) of the lesion, which aims at identifying the lesion and separate it from the background. The algorithm will have to be able to eradicate noise and other undesired features in the image, and to correctly segment the lesion [1]. Visual segmentation of tumor by dermatologist is simple in most of the cases. When transition between lesion and surrounding skin is too smooth, sporadically some irreducible fuzziness remains. The copious papers on boundary detection of skin tumors expound that it is still an open dilemma for computers. As a matter of fact, lesions show a discrepancy in size, color, texture [2].

The process of contour extraction of different objects from background is edge detection and it is very imperative to image understanding and computer vision. Problems with edge detection are edge location errors, false edges, and broken or missing edge fragments [3].

To reduce mortality early detection and surgical excision is currently the only approach, because advanced skin cancers remain incurable. The conventional screening tests require a skin naked-eye examination by an experienced

clinician. ABCD rule is one of the most widely used methods for evaluating pigmented skin lesions with the naked-eye [7]. When the pigmented skin lesions are small or/and regular in shape or color, however, this system may fail sometimes[4]. The most hastily increasing cancer in the world is malignant melanoma. Since melanoma can be cured with a simple expurgation if detected early, early diagnosis is particularly important [5].

Automated border detection is vital for the image analysis because the border structure provides important information for precise diagnosis, as many clinical features such as asymmetry, border irregularity, and abrupt border cutoff are calculated directly from the border.

Automated border detection is a exigent task due to the following reasons: low contrast between the lesion and the surrounding skin, irregular and fuzzy lesion borders, features such as skin lines ,blood vessels , hairs , and air bubbles, variegated coloring inside the lesion , and fragmentation due to various reasons such as scar-like depigmentation [5].

To considerably reduces morbidity and mortality, detection of malignant melanoma should be done in its early stages. We can also hoard hundreds of millions of dollars by early detection that otherwise would be spent on the treatment of advanced diseases. There is a very high likelihood that the patient will survive, if cutaneous melanoma is detected in its early stages and removed. The ABCDs of melanoma are [3]: asymmetry, border irregularity, color variegation, and diameter greater than 6 mm. Image analysis techniques for measuring these features have been developed. Measurement of image features for diagnosis of melanoma requires that first the lesions be detected and localized in an image. It is essential that lesion boundaries are determined accurately so that measurements, e.g. maximum diameter, asymmetry, irregularity of the boundary, and color characteristics can be precisely computed. Various image segmentation methods have been developed for delineating lesion boundaries [6].

II. REVIEW OF RELATED WORK

Due to the great variety of lesions, skin types, presence of hair and so forth, the segmentation stage is not a straightforward task. A variety of image segmentation methods

have been proposed for this purpose. L. Xu et al. developed a three-step segmentation method using the properties of skin cancer images. The steps of their method are as follows: 1. Preprocessing: a color image is first transformed into an intensity image in such a way that the intensity at a pixel shows the color distance of that pixel with the color of the background. The color of the background is taken to be the median color of pixels in small windows in the four corners of the image. 2. Initial segmentation: a threshold value is determined from the average intensity of high gradient pixels in the obtained intensity image. This threshold value is used to find approximate lesion boundaries. 3. Region refinement: a region boundary is refined using edge information in the image. This involves initializing a closed elastic curve at the approximate boundary, and shrinking and expanding it to fit to the edges in its neighbourhood [6].

We have previously developed a segmentation algorithm[17], to extract the true border that reveals the global structure irregularity, which may evoke excessive cell growth or regression of a melanoma. The steps of this algorithm[17] are as follows: 1. This algorithm is applied to the input image containing the lesion, where the input RGB image is converted to grayscale image. 2. Salt and pepper noise is added to the grayscale image and background noise reduction techniques are used to filter noise. 3. The noise filtered image is converted to a binary image, based on threshold. 4. Then the binary image is converted to xor image. 5. The Canny Edge detector is used to find the edges in the xor image. We get the edge detected image. 6. The pixel on the border of the object is found. 7. Using this pixel found on the border of the object (Lesion) as the starting pixel, the border of the lesion is traced, using the segmentation algorithm[17] using canny detector.

Image segmentation is conceivably, the most premeditated area in computer vision, with copious methods reported. A segmentation method is usually designed taking into consideration the properties of a particular class of images. The algorithm will have to be able to confiscate noise and other undesired features in the image, and to correctly segment the lesion. Developing robust and proficient algorithm for medical image segmentation has been a exigent area of interesting research interest, over the last decade [15].

The medical images generally are bound to restrain noise while acquisition. An efficient and robust segmentation algorithm against noise is needed for medical image segmentation. Accurate segmentation of medical images is therefore highly challenging, however, accurate segmentation of these images is imperative in correct diagnosis by clinical tools [16].

In this paper, we have compared the performance of robust segmentation algorithm using log detector for border detection of real skin lesions for noisy skin lesion images developed by us[18], with the segmentation algorithm using canny detector for border detection of real skin lesions for noisy skin lesion images developed by us[17].

III. PROPOSED METHODOLOGY

The image segmentation algorithm using Log Edge Detector, for border detection of skin lesions, developed by us [18], that reveals the global structure irregularity, which may evoke excessive cell growth or regression of a melanoma is discussed in this paper. This algorithm is applied to the image containing the lesion

A. Image Segmentation Algorithm using log edge detector

- Step 1 : The RGB image is converted to grayscale image
- Step 2 : Salt and pepper noise is added to the grayscale image. The noisy image is the input image.
- Step 3: Median filter used as the background noise reduction technique to filter noise.
- Step 4 : After noise reduction, the image is converted to a black and white image, based on threshold,
- Step 5: The black and white image got is converted into xor image
- Step 6: The Log Edge detector is used to find the edges in the xor image. We get the edge detected image.
- Step 7 : The pixel on the border of the object is found. To find the pixel on the border of the object (Lesion) the binary image is used to find the row co-ordinate of the pixel on the border of the object and the edge detected image is used to find the column co-ordinate of the pixel on the border of the object to be traced
- Step 8 : Using this pixel found on the border of the object (Lesion) as the starting pixel, the border of the lesion is traced using the robust segmentation algorithm[18] using log detector, successfully.

B. Median filtering

To reduce "salt and pepper" noise, median filtering is a nonlinear operation often used in image processing. Median filtering is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.

C. Edge Detection

An edge is a set of connected pixels that lie on the boundary between two regions[10]. An image can be segmented by detecting those discontinuities.

The key to a satisfactory segmentation result lies in keeping a balance between detecting accuracy and noise immunity. If the level of detecting accuracy is too high, noise may bring in fake edges making the outline of images unreasonable. Otherwise, some parts of the image outline may get undetected and the position of objects may be mistaken if the degree of noise immunity is excessive [12].

Edge detection is a most common approach for detecting meaningful discontinuities in grey level. Such discontinuities are detected using first order and second order derivatives [5]. The first order derivative of choice is the gradient. The gradient of the 2D function $f(x, y)$, is defined as a vector. The magnitude of this vector is given by

$$g = [Gx^2 + Gy^2]^{1/2} \quad (1)$$

Where $G_x = \partial f / \partial x$ and $G_y = \partial f / \partial y$

The second derivative in image processing is computed using the laplacian. The laplacian is seldom used by itself for edge detection because as a second order derivative it is unacceptably sensitive to noise, its magnitude produces double edges and it is unable to detect edge direction. However Laplacian can be a powerful complement when used in combination of other edge detection techniques. The basic idea behind edge detection is to find places in an image where the intensity changes very rapidly using one of the two general criteria:

1. Find places where the first derivative of the intensity is greater in magnitude than a specified threshold.
2. Find places where the second derivative of the intensity has zero crossing.

1) *Laplacian of Gaussian Detector* : Consider the Gaussian function

$$h(r) = -e^{-r^2/2\sigma^2} \quad (2)$$

Where $r^2 = x^2 + y^2$ and σ is the standard deviation. This is a smoothing function, which if convolved with an image, will blur it. The degree of blurring is determined with the value of σ . The Laplacian of this function (the second derivative with respect to r) is $(-[(r^2 - \sigma^2)/\sigma^4] e^{-r^2/2\sigma^2})$

This function is called Laplacian of Gaussian. Because the second derivative is a linear operation, convolving the image with the above said function, is the same as convolving the image with the smoothing function first and then computing the Laplacian of the result. This is the key concept underlying the LOG detector. The LOG detector finds the edges by looking for zero crossing after filtering $f(x, y)$ with a Gaussian filter [11]

IV. RESULTS AND DISCUSSION

An image segmentation algorithm to extract the true border of the skin lesions, that is helpful in the diagnosis of melanoma, has been implemented using Matlab. Our aim is to select an image and the system should impart an automatic identification (or segmentation) of the lesion, which aims at identifying the lesion and separate it from the background. The algorithm will have to be able to remove noise and other undesired features in the image, and to correctly segment the lesion. The algorithm should work well even when the transition between lesion and surrounding skin is too smooth. The segmentation stage is not a candid task due to the great variety of lesions, skin types, presence of hair etc. The segmentation algorithm using log detector[18], works well even in the presence of noise and hair, to detect the border of the lesion, which helps the medical practitioners in diagnosis.

The robust segmentation algorithm using log detector for border detection of real skin lesions[18] was applied to a variety of skin lesions, and skin types. Figure 1(a), 2(a), 3(a), 4(a), 5(a), 6(a), 7(a), 8(a) and 9(a) illustrates different types of original skin lesions. Figure 1(c), 2(c), 3(c), 4(c), 5(c), 6(c), 7(c), 8(c) and 9(c) shows the final output results of the segmentation algorithm using canny edge detector[17], by

J.H.Jaseema Yasmin et al. [17], when they are applied to the different types of skin lesions, with noise.

For the different types of Skin Lesions taken, J.H.Jaseema Yasmin et al. [17] method poorly delineates the boundary for some of the skin lesions. The Figure 1(c), 2(c), 3(c) demonstrates the failure of this method[17] to delineate the boundary of the lesion of various types.

The robust segmentation algorithm using log detector[18], converts the original skin lesion image (skin lesion 1-9) in Figure 1(a), 2(a), 3(a), 4(a), 5(a), 6(a), 7(a), 8(a) and 9(a) into a gray scale image. 20% salt and pepper noise was added to the original image and that is illustrated in Figure 1(b), 2(b), 3(b), 4(b), 5(b), 6(b), 7(b), 8(b) and 9(b). The noisy image is the input image to the proposed algorithm. The median filter is applied and the noise is removed. After noise removal the image is enhanced. Based on a threshold value the enhanced image is converted to black and white image. This algorithm[18] converts the black and white image into xor image and the edges are detected using log edge detector. The black and white image is used to find the row co-ordinate of the pixel on the border of the object and the edge detected image is used to find the column co-ordinate of the pixel on the border of the object to be traced and using this pixel found on the border of the object as the starting pixel, the border of the lesion is traced using the robust segmentation algorithm[18] successfully is shown in Figure 1(d), 2(d), 3(d), 4(d), 5(d), 6(d), 7(d), 8(d) and 9(d). The robust segmentation algorithm using log detector[18], segments the lesion from the image even in the presence of noise and presence of hair for a variety of lesions, and skin types.

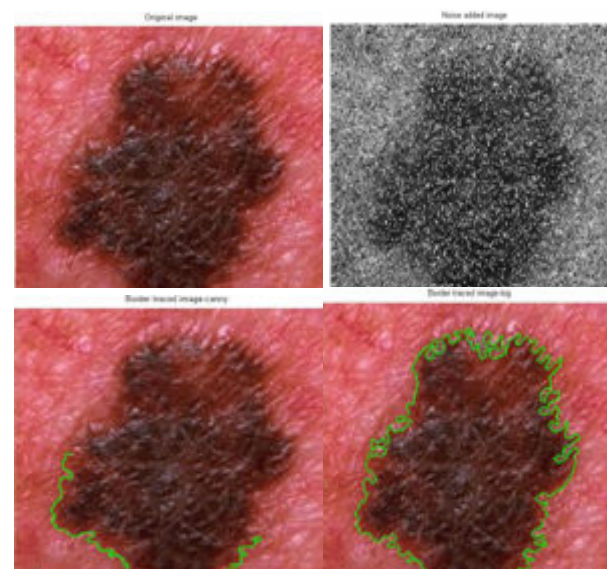


Figure 1. Demonstration of border detection for Skin lesion 1
(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.

Figure 1(a) referred from L.Xua et.al [6]

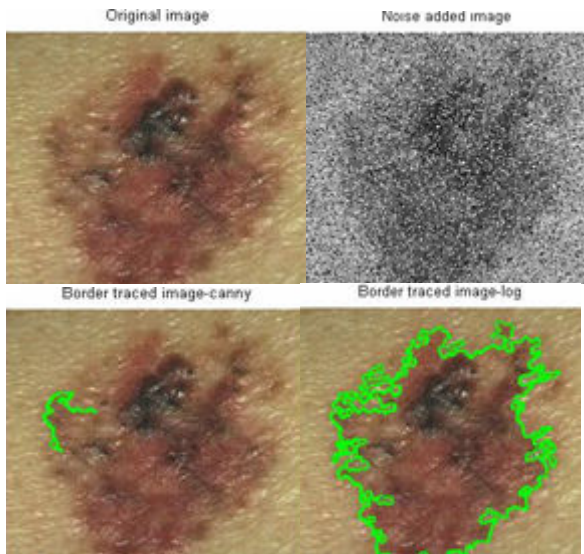


Figure 2.Demonstration of border detection for Skin lesion 2
(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.
Figure 2(a) referred from L.Xua et.al [6]

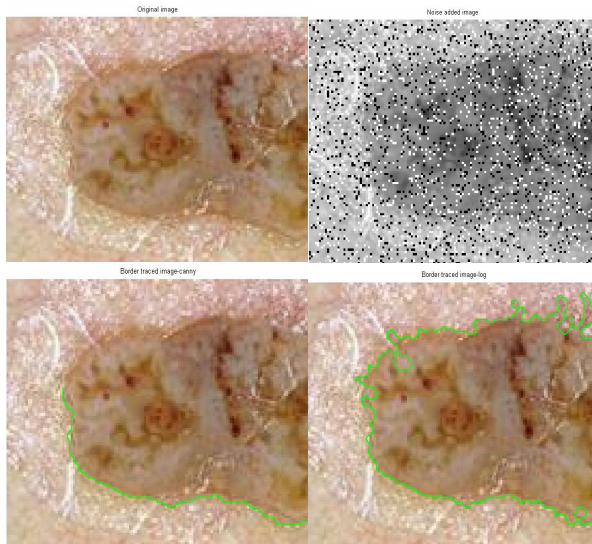


Figure 3.Demonstration of border detection for Skin lesion 2
(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.
Figure 3(a) referred from M.Emre Celebia et.al [5]

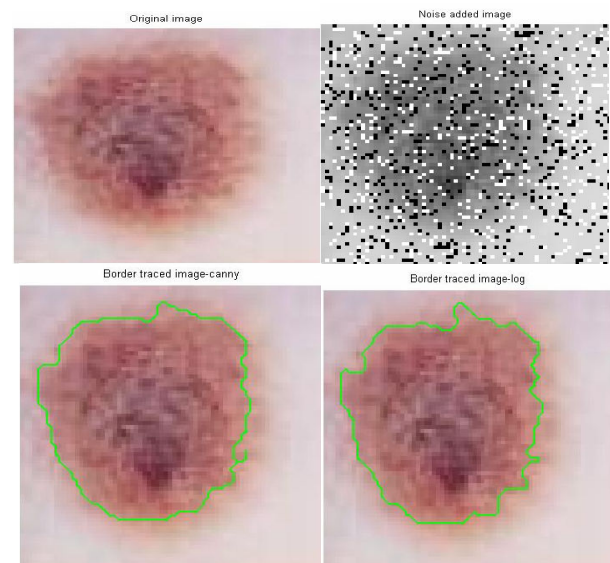


Figure 4.Demonstration of border detection for Skin lesion 4
(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.
Figure 4(a) referred from L.Xua et.al [6]

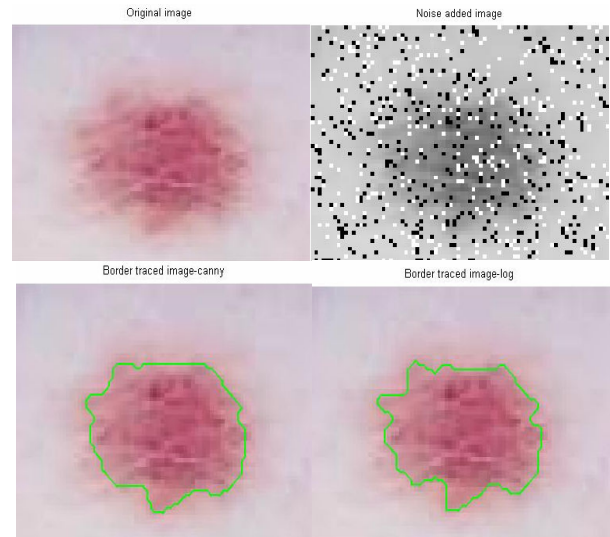


Figure 5.Demonstration of border detection for Skin lesion 5
(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.
Figure 5(a) referred from L.Xua et.al [6]

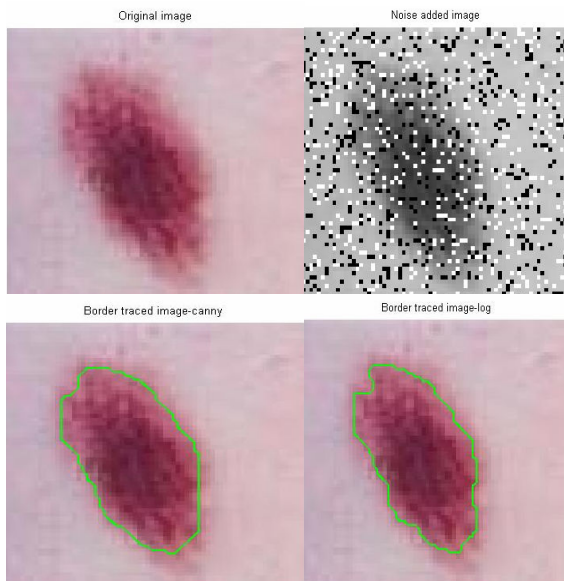


Figure 6.Demonstration of border detection for Skin lesion 6

(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.

Figure 6(a) referred from L.Xua et.al [6]

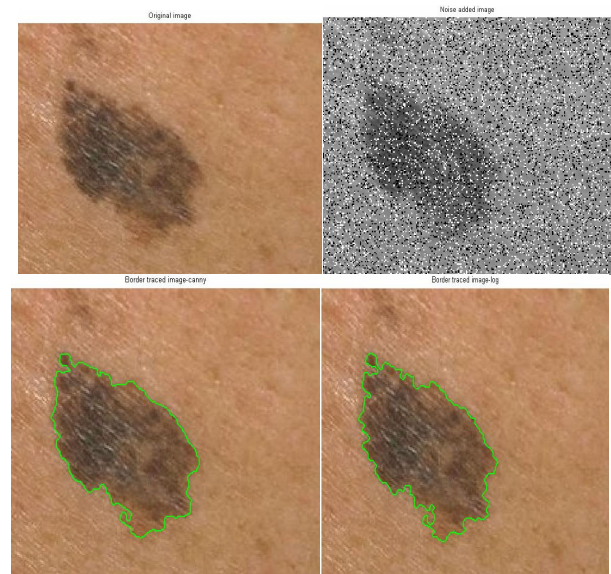


Figure 8.Demonstration of border detection for Skin lesion 8

(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.

Figure 8(a) referred from L.Xua et.al [6]

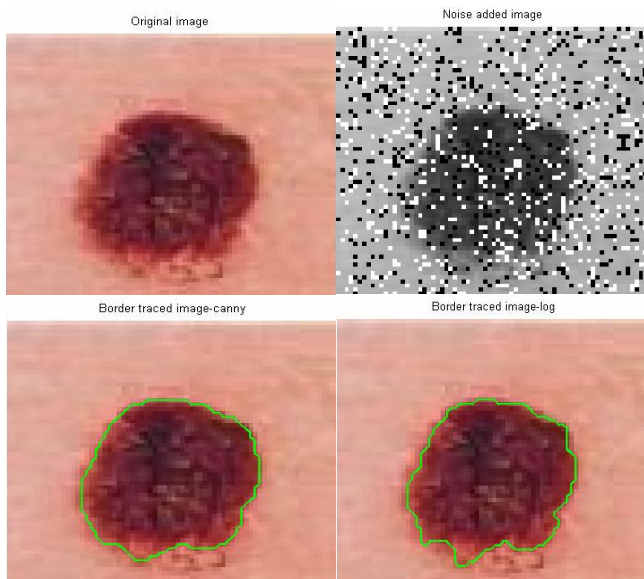


Figure 7.Demonstration of border detection for Skin lesion 7

(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.

Figure 7(a) referred from L.Xua et.al [6]

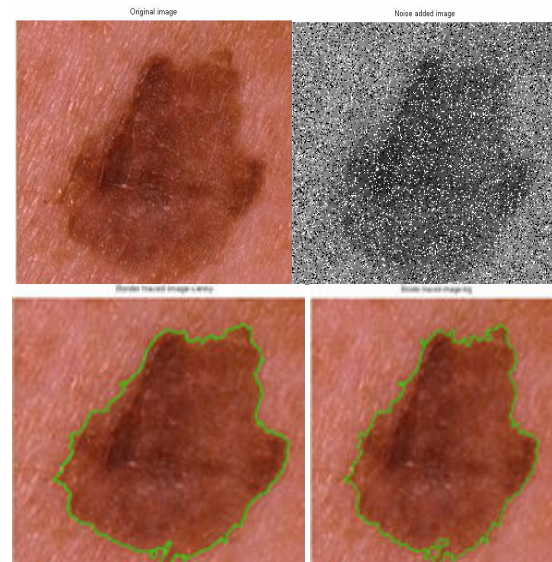


Figure 9.Demonstration of border detection for Skin lesion 9

(a) Skin lesion (b) Noisy image (c) Border traced image by robust segmentation algorithm using canny detector[17] for a noisy image. (d) Border traced image by robust segmentation algorithm using LOG detector for a noisy image.

Figure 9(a) referred from L.Xua et.al [6]

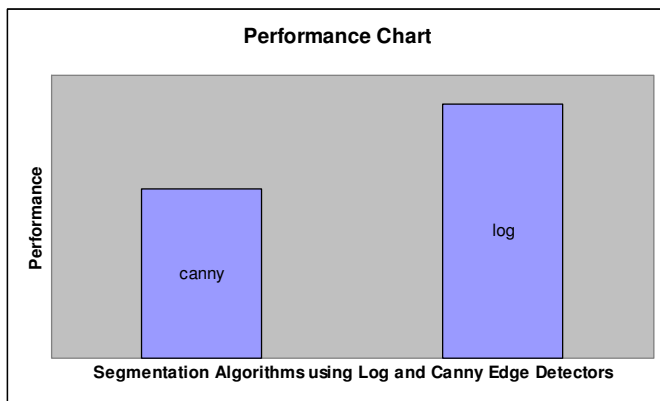


Figure10. Performance comparison chart of the segmentation algorithm using LOG edge detector and Canny edge detector for tracing the border of noisy skin lesion images

The Performance comparison chart of the segmentation algorithm using log edge detector[18] and Canny edge detector[17] for tracing the border of noisy skin lesion images is shown in Figure 10. The segmentation algorithm which uses canny detector to trace the edges of the skin lesion in a noisy skin image fail to detect the edges in some of the cases as shown in Figure 1(c), Figure 2(c), Figure 3(c). The segmentation algorithm which uses log detector to trace the edges of the skin lesion in a noisy skin image successfully detects the edges in all of the cases as shown in Figure 1(d) to Figure 9(d). So the performance of the segmentation algorithm using log edge detector for tracing the border of noisy skin lesion images is better than the performance of the segmentation algorithm using canny edge detector for tracing the border of noisy skin lesion images.

V. Conclusion

In this paper, we have discussed about the effective segmentation algorithm using log edge detector[18], and about the segmentation algorithm using canny detector[17], both developed by us, for border detection of real skin lesions and compared their performance in the border detection of real skin lesions. To verify the capability of the segmentation algorithm in detecting the border of the lesions for skin lesion diagnosis, the algorithm was applied on diversity of clinical skin image containing lesions with noise. The experimental results demonstrated the successful border detection of real skin lesions by the segmentation algorithm using log detector[18] for clinical skin images with noise and make them accessible for further analysis and research. We conclude that the segmentation algorithm using log detector[18] segments the lesion from the image even in the presence of noise and presence of hair for a variety of lesions, and skin types and we conclude that its performance is better than the segmentation algorithm that uses canny detector[17], for border detection of real skin lesions for noisy skin lesion diagnosis.

REFERENCES

- [1] Teresa Mendonça, André R. S. Marc, Angela Vieira Jacinto C. Nascimento, Margarida Silveira, Jorge S. Marques, Jorge Rozeira, "Comparison of Segmentation Methods for Automatic Diagnosis of Dermoscopy Images", Proc. of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France, August 23-26, 2007, pp. 6572-6575
- [2] Arthur Tenenhaus, Alex Nkengne, Jean-François Horn, Camille Serruys, Alain Giron, and Bernard Fertil, "Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions", Author manuscript, published in Skin Research and Technology Vol. 16, N° 1 (2009) pp. 85-97
- [3] LI Xue-wei, ZHANG Xin-rong, "A Perceptual Color Edge Detection Algorithm", 2008 International Conference on Computer Science and Software Engineering, pp. 297-300
- [4] Alfonso Baldi, Marco Quartulli, Raffaele Murace, Emanuele Dragonetti, Mario Manganaro, Oscar Guerra, and Stefano Bizzi, Automated Dermoscopy Image Analysis of Pigmented Skin Lesions, Cancers 2010, 2, 262-273
- [5] M. Emre Celebi, Hitoshi Iyatomi, Gerald Schaefer, William V. Stoecker, "Lesion border detection in dermoscopy images", Computerized Medical Imaging and Graphics 33 (2009) 148-153
- [6] L. Xu, M. Jackowska, A. Goshtasby, D. Roseman, S. Binesh, C. Yuc, A. Dhawand, A. Huntley, "Segmentation of skin cancer images", Image and Vision Computing 17 (1999) 65-74
- [7] Harald Ganster, Axel Pinz, Reinhard Röhner, Ernst Wildling, Michael Binder, and Harald Kittler, "Automated Melanoma Recognition", IEEE Transactions on Medical Imaging, vol. 20, no. 3, March 2001
- [8] Roberto Rodríguez and Ana G. Suarez, "A new algorithm for image segmentation by using iteratively the mean shift filtering", Scientific Research and Essay Vol. 1 (2), pp. 043-048, November 2006
- [9] Wen-Xiong Kang, Qing-Qiang Yang, Run-peng Liang, "The Comparative Research on Image Segmentation Algorithms", 2009 First International Workshop on Education Technology and Computer Science.
- [10] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", second edition, Prentice-Hall, India.
- [11] Rafael C. Gonzalez, Richard E. Woods, Stevan L. Eddins "Digital Image Processing using Matlab"
- [12] Wen-Xiong Kang, Qing-Qiang Yang, Run-peng Liang, "The Comparative Research on Image Segmentation Algorithms", 2009 First International Workshop on Education Technology and Computer Science.
- [13] Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D. Fox, "Level Set Evolution Without Re-initialization: A New Variational Formulation", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
- [14] J.H. Jaseema Yasmin, M. Mohamed Sathik, S. Zulaikha Beevi, "Edge Detection Algorithms for Medical Image Segmentation", Proceedings of the International Conference on Intelligent Design and Analysis of Engineering Products Systems and Computation 9-10 July 2010 (IDAPSC-10), Coimbatore, Pg 63.

- [15] S.Zulaikha Beevi,M.Mohamed Sathik, "A Robust segmentation Approach for Noisy Medical Images Using Fuzzy Clustering with Spatial Probability" ,European Journal of Scientific Research,Vol. 41,No.3(2010),pp.437-451
- [16] S.Zulaikha Beevi,M.Mohamed Sathik, K.Senthamarai Kannan ,J.H.Jaseema Yasmin, "Hybrid Segmentation Approach Using FCM and Dominant Intensity Grouping with Region Growing on Medical image"International Journal of Advanced Research in Computer Science, Vol. 1, No. 2, July-August 2010,pp.103-108 (Review Article)
- [17] J.H.Jaseema Yasmin, M.Mohamed Sathik, S.Zulaikha Beevi, "Effective Border Detection of Noisy Real Skin Lesions for Skin Lesion Diagnosis by Robust Segmentation Algorithm",International Journal of Advanced Research in Computer Science,Vol. 1, No. 3, Sept-october 2010,pp.110- 116.(Research paper)
- [18] J.H.Jaseema Yasmin, M.Mohamed Sathik, S.Zulaikha Beevi, "Robust Segmentation Algorithm using LOG Edge Detector for Effective Border Detection of Noisy Skin Lesions", ICCET 2011, pp. 61-66.

Query Data With Fuzzy Information In Object-Oriented Databases An Approach The Semantic Neighborhood Of Hedge Algebras

Doan Van Thang

Korea-VietNam Friendship Information Technology College
Department of Information systems, Faculty of Computer
Science Da Nang City, Viet Nam
vanthangdn@gmail.com

Doan Van Ban

Institute of Information Technology, Academy Science and
Technology of Viet Nam.
Ha Noi City, Viet Nam

Abstract - In this paper, we present an approach for handling attribute values of object classes with fuzzy information and uncertainty in object-oriented database based on theory hedge algebraic. In this approach, semantics be quantified by quantitative semantic mapping of hedge algebraic that still preserving in order semantics may allow manipulation data on the real domain of attribute in relation with the semantics of linguistic. And then, evaluating semantics, searching information uncertainty, fuzziness and classical data entirely consistent based on the ensuring homogeneity of data types. Hence, we present algorithm that allow the data matching helping the requirements of the query data.

I. INTRODUCTION

In approach interval value [2], we consider to attributive values object class is interval values and the interval values are converted into sub interval in $[0, 1]$ respectively and then we perform matching interval this. However, attributive value of the object in the fuzzy object-oriented database is complex: linguistic values, reference to objects (this object may be fuzzy), collections,... Thus matching data also become more complex. Hence, query information method proposed in [2] is not satisfy requirements for the case of this data yet.

In this paper, we research has expanded for handling attribute value is linguistic value. There are many approaches on handling fuzzy information with linguistic semantic that researchers interests [1], [3]. We based on approach hedge algebra, where linguistic semantic is obtained by considering the terms as expressed by the part of order relation. In this approach linguistic value is data which is not label of fuzzy set representation semantic of linguistic value. Using quantitative semantics mapping of hedge algebra to transfer linguistic values into real values that preserve in order semantics may allow manipulation data on the real domain of attribute in relation with the semantics of linguistic.

The paper is organized as follows: Section 2 presenting the basic concepts relevant to hedge algebraic as the basis for the next sections; section 3 proposing SASN (Search Attributes in the Semantic Neighborhood) and SMSN (Search Method in the Semantic Neighborhood) algorithms for searching data fuzzy conditions for both attributes and methods; section 4 presenting examples for searching data with fuzzy information, and finally conclusion.

II. FUNDAMENTAL CONCEPTS

In this section, we present some fundamental concepts related to hedge algebra [5].

Let hedge algebra $\underline{X} = (X, G, H, \leq)$, where $X = LDom(\underline{X})$, $G = \{1, c^-, W, c^+, 0\}$ is set generator terms, H is a set of hedge considered as a one-argument operations and \leq relation on terms (fuzzy concepts) is a relation order "induced" from natural semantics on X . Set X is generated from G by means of one-argument operations in H . Thus, a term of X represented as $x = h_n h_{n-1} \dots h_1 x$, $x \in G$. Set of terms is generated from the an X term denoted by $H(x)$. Let set hedges $H = H^- \cup H^+$, where $H^+ = \{h_1, \dots, h_p\}$ and $H^- = \{h_{-1}, \dots, h_{-q}\}$ are linearly ordered, with $h_1 < \dots < h_p$ and $h_{-1} < \dots < h_{-q}$, where $p, q > 1$, we have the following definitions related:

Definition 2.1 An $fm : X \rightarrow [0, 1]$ is said to be a fuzziness measure of terms in X if:

- (1) fm is called complete, that is $\forall u \in X$,
$$\sum_{-q \leq i \leq p, i \neq 0} fm(h_i u) = fm(u)$$
- (2) if x is precise, that is $H(x) = \{x\}$ then $fm(x) = 0$.
Hence $fm(0) = fm(W) = fm(1) = 0$.

(3) $\forall x, y \in X, \forall h \in H, \frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$, This proportion is called the fuzziness measure of the hedge h and denoted by $\mu(h)$.

Definition 2.2 (Quantitative semantics function v)

Let fm is fuzziness measure of X , quantitative semantics function v on X is defined as follows:

- (1) $v(W) = \theta = fm(c^-)$, $v(c^-) = \theta - \alpha fm(c^-)$ and $v(c^+) = \theta + \alpha fm(c^+)$
- (2) If $1 \leq j \leq p$ then:
$$v(h_j x) = v(x) + Sign(h_j x) \times \left[\sum_{i=1}^j fm(h_i x) - \omega(h_j x) fm(h_j x) \right]$$
- (3) If $-q \leq j \leq -1$ then:
$$v(h_j x) = v(x) + Sign(h_j x) \times \left[\sum_{i=j}^{-1} fm(h_i x) - \omega(h_j x) fm(h_j x) \right]$$

Where:

$$\omega(h_j x) = \frac{1}{2} [1 + \text{Sign}(h_j x) \text{Sign}(h_q h_j x) (\beta - \alpha)] \in \{\alpha, \beta\}$$

Definition 2.3 Invoke fm is fuzziness measure of hedge algebra X , $f: X \rightarrow [0, 1]$, $\forall x \in X$, denoted by $I(x) \subseteq [0, 1]$ and $|I(x)|$ is measure length of $I(x)$.

A family $J = \{I(x): x \in X\}$ called the partition of $[0, 1]$ if:

(1): $\{I(c^+), I(c^-)\}$ is partition of $[0, 1]$ so that $|I(c)| = fm(c)$, where $c \in \{c^+, c^-\}$.

(2): If $I(x)$ defined and $|I(x)| = fm(x)$ then $\{I(h_i x): I = 1 \dots p+q\}$ is defined as a partition of $I(x)$ so that satisfy conditions: $|I(h_i x)| = fm(h_i x)$ and $|I(h_i x)|$ is linear ordering.

Set $\{I(h_i x)\}$ called the partition associated with the terms x . We have

$$\sum_{i=1}^{p+q} |I(h_i x)| = |I(x)| = fm(x)$$

Definition 2.4 Set $X_k = \{x \in X : |x| = k\}$, consider $P^k = \{I(x) : x \in X_k\}$ is a partition of $[0, 1]$. Its said that u equal v at k level, denoted by $u \equiv_k v$, if and only if $I(u)$ and $I(v)$ together included in fuzzy interval k level. Denote $\forall u, v \in X$, $u \equiv_k v \Leftrightarrow \exists \Delta^k \in P^k : I(u) \subseteq \Delta^k$ and $I(v) \subseteq \Delta^k$.

III. DATA SEARCH METHOD

Let fuzzy class $C = (\{a_1, a_2, \dots, a_n\}, \{M_1, M_2, \dots, M_m\})$; o is object of fuzzy class C . Denoted $o.a_i$ is attribute value of o on attribute a_i ($1 \leq i \leq n$) and $o.M_j$ is value method of o ($1 \leq j \leq m$).

In [2] we presented the attribute values are 4 cases: precise value; imprecise value (or fuzzy); object; collection. In this paper, we only interested in handing case 1 and 2: precise value and imprecise value (fuzzy value) and to see precise value is particular case of fuzzy value. Fuzzy value is complex and linguistic label is often used to represent the value of this type. Domain fuzzy attribute value is the union two components:

$$\text{Dom}(a_i) = \text{CDom}(a_i) \cup \text{FDom}(a_i) \quad (1 \leq i \leq n).$$

Where:

- $\text{CDom}(a_i)$: domain crisp values of attribute a_i .
- $\text{FDom}(a_i)$: domain fuzzy values of attribute a_i .

A. Neighborhood level k

We can get fuzzy interval of terms length k as the similarity between terms. It means that the term that representative value of them depending on fuzzy interval level k is similar level k . However, to build the fuzzy interval level k , representative value of terms x have length less than k is always in the end of fuzzy interval level k . Hence, when determining neighborhood level k , we expect representative value it must be inner point of neighborhood level k .

Based on fuzzy interval level k and $k+1$ we construct a partition of the domain $[0, 1]$ following as [8]:

(1) *Similar level 1*: with $k = 1$, fuzzy interval level 1 including $I(c^-)$ and $I(c^+)$. fuzzy interval level 2 on interval $I(c^+)$ is $I(h_{-q} c^+) \leq I(h_{-q+1} c^+) \dots \leq I(h_{-2} c^+) \leq I(h_{-1} c^+) \leq v_A(c^+) \leq I(h_1 c^+) \leq I(h_2 c^+) \leq \dots \leq I(h_{p-1} c^+) \leq I(h_p c^+)$. Meanwhile, we construct partition at similar level 1 include the equivalence classes following: $S(\emptyset) = I(h_p c^-)$; $S(c^-) = I(c^-) \setminus [I(h_{-q} c^-) \cup I(h_p c^-)]$; $S(W) = I(h_{-q} c^-) \cup I(h_{-q} c^+)$; $S(c^+) = I(c^+) \setminus [I(h_{-q} c^+) \cup I(h_p c^+)]$ and $S(I) = I(h_p c^+)$.

We see that except the two end points $v_A(\emptyset) = 0$ and $v_A(I) = 1$, representative values $v_A(c^-)$, $v_A(W)$ and $v_A(c^+)$ are inner point corresponding of classes similar level 1 $S(c^-)$, $S(W)$ and $S(c^+)$.

(2) *Similar level 2*: with $k = 2$, fuzzy interval level 2 including $I(h_i c^+)$ and $I(h_i c^-)$ with $-q \leq i \leq p$. We have equivalence classes following: $S(\emptyset) = I(h_p h_p c^-)$; $S(h_i c^-) = I(h_i c^-) \setminus [I(h_{-q} h_i c^-) \cup I(h_p h_i c^-)]$; $S(W) = I(h_{-q} h_{-q} c^-) \cup I(h_{-q} h_{-q} c^+)$; $S(h_i c^+) = I(h_i c^+) \setminus [I(h_{-q} h_i c^+) \cup I(h_p h_i c^+)]$ and $S(I) = I(h_p h_p c^+)$, with $-q \leq i \leq p$.

By the same, we can construct partition equivalence classes level k at any. However, in fact, $k \leq 4$ and it means that there is maximum 4 hedges consecutive action onto primary terms c^- and c^+ . Precise and fuzzy values will be at the similar level k if the representative value of their in the same class similar level k .

Hence, neighborhood level k of fuzzy concept is determining following: Assuming partition the class similar level k is intervals $S(x_1), S(x_2), \dots, S(x_m)$. Meanwhile, every fuzzy value fu is only and only belong to a similar class. Instance for $S(x_i)$ and called neighborhood level k of fu and denoted by $FRN_k(fu)$.

B. Relation matching on domain of fuzzy attribute value

Based on the concept neighborhood, we give the definition of the relation matching between terms in the domain of the fuzzy attribute value.

Definition 3.1

Let fuzzy class C determine on the set of attributes A and methods M , $a_i \subseteq A$, $o_1, o_2 \in C$. We say that $o_1.a_i \equiv_k o_2.a_i$ and equal level k if:

- (1) If $o_1.a_i, o_2.a_i \in \text{CDom}(a_i)$ then $o_1.a_i = o_2.a_i$ or existence $FRN_k(x)$ such that $o_1.a_i, o_2.a_i \in FRN_k(x)$.
- (2) If $o_1.a_i$ or $o_2.a_i \in \text{FDom}(a_i)$, instance for $o_1.a_i$ then we have to $o_2.a_i \in FRN_k(o_1.a_i)$.
- (3) If $o_1.a_i, o_2.a_i \in \text{FDom}(a_i)$ then $FRN_k(o_2.a_i) = FRN_k(o_1.a_i)$.

Definition 3.2

Let fuzzy class C determine on the set of attributes A and methods M , $a_i \subseteq A$, $o_1, o_2 \in C$. We say that $o_1.a_i \geq_k o_2.a_i$ if:

- (1) If $o_1.a_i, o_2.a_i \in \text{CDom}(a_i)$ then $o_1.a_i \geq o_2.a_i$.
- (2) If $o_1.a_i$ and $o_2.a_i \in \text{FDom}(a_i)$ then we have to

- $o_1.a_i \geq FRN_k(o_2.a_i)$.
(3) If $o_1.a_i, o_2.a_i \in FDom(a_i)$ then
 $FRN_k(o_1.a_i) \geq FRN_k(o_2.a_i)$

C. Algorithm search data approach to semantic neighborhood

In [2] we presented the structure of fuzzy OQL queries are considered as: *select <attributes>/<methods> from <class> where <fc>*, where <fc> are fuzzy conditions or combination of fuzzy condition that allow using of disjunction or conjunction operations.

In this paper, we use approaching to semantic neighborhood for determinating the truth value of the <fc> and associated truth values.

Example, we consider query following “show all students are possibly young age”. To answer this query, we perform following:

+ Step 1: We construct intervals similar level k , $k \leq 4$ because it's a maximum 4 hedges consecutive action onto primary terms c^- and c^+ .

+ Step 2: Determine neighborhood level k of fuzzy condition. In the above query, fuzzy condition is possibly young should neighborhood level 2 of possibly young is $FRN_2(\text{possibly young})$, and determine neighborhood level 2 of fuzzy attribute value is $FRNA_2(attr)$. At last based on definition 3.1, we perform data matching two neighborhood level 2 of $FRNA_2(attr)$ and $FRN_2(\text{possibly young})$.

Without loss of generality, we consider on cases multiple fuzzy conditions with notation follow as:

- \mathcal{G} is AND or OR operation.
- $fzvalue_i$ is fuzzy values of the i attribute.

On that basis, we built the SASN algorithms

SASN algorithm: search data in cases multiple fuzzy conditions for attribute with operation \mathcal{G} .

Input: A class $C = (\{a_1, a_2, \dots, a_n\}, \{M_1, M_2, \dots, M_m\})$, $C = \{o_1, o_2, \dots, o_n\}$.

where $a_i, i = 1 \dots p$ is attribute, M_j is methods.

Output: Set of objects $o \in C$ satisfy condition $\bigcap_{i=1}^p (o.a_i = fzvalue_i)$.

Method

// Initialization.

- (1) **For** $i = 1$ **to** p **do**
- (2) **Begin**
- (3) Set $G_{a_i} = \{\emptyset, c_{a_i}^-, W, c_{a_i}^+, I\}$;
 $H_{a_i} = H_{a_i}^+ \cup H_{a_i}^-$. Where $H_{a_i}^+ = \{h_1, h_2\}$, $H_{a_i}^- = \{h_3, h_4\}$, with $h_1 < h_2$ and $h_3 > h_4$. Select the fuzzy measure for the generating term and hedge.
- (4) $D_{a_i} = [\min_{a_i}, \max_{a_i}]$ // \min_{a_i}, \max_{a_i} : min and max value of domain a_i .
- (5) $FD_{a_i} = H_{a_i}(c_{a_i}^+) \cup H_{a_i}(c_{a_i}^-)$.
- (6) **End**

- (7) Determine intervals level k of fuzzy condition: kQ .

// Partition D_{a_i} into interval similar level k .

- (8) $k = kQ$; // level partition largest with $k = 4$

- (9) **For** $i = 1$ **to** p **do**

- (10) **For** $j = 1$ **to** $2^5(k-1)$ **do**

- (11) Construct intervals similar level k : $S_{a_i}^k(x_j)$;

// Determine neighborhood level k of $o.a_i$.

- (12) **For** each $o \in C$ **do**

- (13) **For** $i = 1$ **to** p **do**

- (14) **Begin**

- (15) $t=0$;

- (16) **Repeat**

- (17) $t=t+1$;

- (18) **Until** $o.a_i \in S_{a_i}^k(x_i)$ or $t > 2^5(k-1)$;

- (19) $FRNA_i^k(attr_i) = FRNA_i^k(attr_i) \cup S_{a_i}^k(x_i)$;

- (20) **End**

// Determine neighborhood level k of $fzvalue_i$.

- (21) **For** $i = 1$ **to** p **do**

- (22) **Begin**

- (23) $t=0$;

- (24) **Repeat**

- (25) $t=t+1$;

- (26) **Until** $fzvalue_i \in S_{a_i}^k(x_i)$ or $t > 2^5(k-1)$;

- (27) $FRN_i^k(fzvalue_i) = FRN_i^k(fzvalue_i) \cup S_{a_i}^k(x_i)$;

- (28) **End**

- (29) $result = \emptyset$;

- (30) **For** each $o \in C$ **do**

- (31) **if** $\bigcap_{i=1}^p (FRNA_i^k(attr_i) = FRNA_i^k(fzvalue_i))$
then $result = result \cup \{o\}$;

- (32) **Return** $result$;

Similar to the method we have SMSN algorithm following:

SMSN algorithm: search data cases single fuzzy conditions for method.

Search data in this case, the first we determine neighborhood level k fuzzy conditions of method is $FRNP_k(fzpvalue)$. Further, we determine neighborhood level k of attributes which method handling: $FRNA_k(attr1)$, $FRNA_k(attr2)$, ..., $FRNA_k(attrn)$. We choose the function combination of hedge algebras being consistent with method that it operate. Then, neighborhood level k of function combination is $FRNPA_k(x)$.

At last based on definition 3.1, we perform data matching two neighborhood level k of $FRNP_k(fzpvalue)$ and $FRNPA_k(x)$.

Input: A class $C = (\{a_1, a_2, \dots, a_n\}, \{M_1, M_2, \dots, M_m\})$, $C = \{o_1, o_2, \dots, o_n\}$.

where $a_i, i = 1 \dots p$ is attribute, M_j is methods.

Output: Set of objects $o \in C$ satisfy condition $(o.M_i = fzpvalue)$.

Method

```

// Initialization.
(1) For i = 1 to p do
(2) Begin
(3) Set  $G_{a_i} = \{ \emptyset, c_{a_i}^-, W, c_{a_i}^+, I \}$ ;  $H_{a_i} = H_{a_i}^+ \cup H_{a_i}^-$ .
    Where  $H_{a_i}^+ = \{h_1, h_2\}$ ,  $H_{a_i}^- = \{h_3, h_4\}$ , with  $h_1 < h_2$ 
    and  $h_3 > h_4$ . Select the fuzzy measure for the
    generating term and hedge.
(4)  $D_{a_i} = [\min_{a_i}, \max_{a_i}]$  //  $\min_{a_i}, \max_{a_i}$ : min and
    max value of domain  $a_i$ .
(5)  $FD_{a_i} = H_{a_i}(c_{a_i}^+) \cup H_{a_i}(c_{a_i}^-)$ .
(6) End
(7) Determine intervals level  $k$  of fuzzy condition:  $kQ$ .
//Partition  $D_{a_i}$  into interval similar level  $k$ .
(8)  $k = kQ$ ; // level partition largest with  $k = 4$ 
(9) For i = 1 to p do
(10) For j = 1 to  $2^5(k-1)$  do
(11) Construct intervals similar level  $k$ :  $S_{a_i}^k(x_j)$ ;
//Determine neighborhood level  $k$  of  $o.a_i$ .
(12) For each  $o \in C$  do
(13) For i = 1 to p do
(14) Begin
(15)  $t=0$ ;
(16) Repeat
(17)  $t=t+1$ ;
(18) Until  $o.a_i \in S_{a_i}^k(x_i)$  or  $t > 2^5(k-1)$ ;
(19)  $FRNA_i^k(attr_i) = FRNA_i^k(attr_i) \cup S_{a_i}^k(x_i)$ ;
(20) End
// Determine neighborhood level  $k$  of  $fzpv$  value.
(21)  $i = 1$ ;  $f = 0$ ;
(22) While ( $i \leq p$ ) and ( $f = 0$ ) do
(23) Begin
(24)  $j=0$ ;
(25) While ( $j \leq 2^5(k-1)$ ) and ( $f = 0$ ) do
(26) Begin
(27)  $j=j+1$ ;
(28) if  $fzpv$  value  $\in S_{a_i}^k(x_j)$  then  $f = 1$ ;
(29) End;
(30)  $i = i + 1$ ;
(31) End
(32)  $FRNP^k(fzpv) = S_{a_i}^k(x_j)$ ;
(33) For each  $o \in C$  do
(34) For  $i=1$  to  $m$  do
(35) function combination hedge algebras:

$$FRNPA_i^k(x_i) = \bigotimes_{j=1}^p (FRNA_j^k(attr_j));$$

(36)  $result = \emptyset$ ;
//Combination of hedge algebras with operation  $\mathcal{G}$  is
operation and
(37) For each  $o \in C$  do
(38) For  $i=1$  to  $m$  do
(39) if  $FRNPA_i^k(x_i) = FRNP^k(fzpv)$ 

```

then $result = result \cup \{o\}$;

(40) Return result;

Theorem: SASN algorithm and SMSN algorithm always stop and correct.

Proof:

1. *The Stationarity:* Set of attributes, the method of the object is finite (n, p, m is finite) so algorithm will stop when all objects completed the approved.

2. *The corrective maintenance:*

Really, for each attribute a_i ($1 \leq i \leq n$) in object $o \in C$, the attribute values can get a classic value (precise value) or linguistic value (fuzzy value). In relation matching for data, we are divided into the following two cases:

First case: For classic attribute values (precise value), we use operation $=$ to perform data matching.

Second case: For linguistic value, we use operation matching at level $=_k$, with k is interval neighborhood level k by hedge algebra. Based on quantitative semantics, we determined neighborhood level k of term x is $FRN_k(x) = [a, b]$, the following cases:

a) If y is classic value (precise value) that $y \in [a, b]$ then $y =_k x$.

b) If y is linguistic value in interval $[x_1, x_2]$ (it is calculated through quantitative semantics) that $a \leq x_1$ and $x_2 \leq b$ then $y =_k x$.

Two algorithms are implemented to matching data in case data is classical or linguistic values and the output is corrective.

Computational complexity of SASN algorithm evaluation follows as: step (1) - (19) complexity is $O(p)$, step (20) - (32) is $O(n*p)$. So, the SASN algorithm can computational complexity $O(n*p)$.

Computational complexity of SMSN algorithm evaluation follows as: step (1) - (23) complexity is $O(p)$, step (24) - (32) is $O(n*p)$, step (33) - (36) is $O(m*n*p)$, step (37) - (40) is $O(m*n)$. So, the SMSN algorithm can computational complexity $O(n*p*m)$.

IV. EXAMPLE

We consider a database with six rectangular objects as follows:

Rectangular				
iD	name	length of edges	width of edges	area()
iD1	hcn1	62	Little short	
iD2	hcn2	53	55.5	
iD3	hcn3	very very short	70	
iD4	hcn4	58	very long	
iD5	hcn5	little long	45	
iD6	hcn6	55	Little short	

Query 1: List of rectangles have length "Little long" or width "Little short"

Using algorithms SASN the following:

Step (1) - (6):

Let consider a linear hedge algebra of length, $X_{length} = (X_{length}, G_{length}, H_{length}, \leq)$, where $G_{length} = \{short, long\}$, $H_{length}^+ = \{More, Very\}$, $H_{length}^- = \{Possibly,$

Little}, where *P*, *L*, *M* and *V* stand for *Possibly*, *Little*, *More* and *Very*, with *Very* > *More* and *Little* > *Possibly*.

Suppose that $W_{length} = 0.6$, $fm(short) = 0.6$, $fm(long) = 0.4$, $fm(V) = 0.35$, $fm(M) = 0.25$, $fm(P) = 0.2$, $fm(L) = 0.2$.

$Dom(DODAI) = [0, 100]$. $Result = \emptyset$;
 $LD_{length} = H_{length}(short) \cup H_{length}(long)$.

Step (7) - (20): so *little long* and *little short* = 2 so we only need to build interval *similar level 2*. We perform partition the interval [0, 100] into interval *similar level 2*:

$fm(VVshort) = 0.35 * 0.35 * 0.6 * 100 = 7.35$, so $S(0) = [0, 7.35]$;

$fm(MVshort) + fm(PVshort) = (0.25 * 0.35 * 0.6 + 0.2 * 0.35 * 0.6) * 100 = 9.45$, so $S(Vshort) = (7.35, 16.8]$;

$fm(LVshort) + fm(VMshort) = (0.2 * 0.35 * 0.6 + 0.35 * 0.25 * 0.6) * 100 = 9.45$; $fm(MMshort) + fm(PMshort) = (0.25 * 0.25 * 0.6 + 0.2 * 0.25 * 0.6) * 100 = 6.75$, so $S(Mshort) = (26.25, 33]$;

$fm(LMshort) + fm(VPshort) = (0.2 * 0.25 * 0.6 + 0.35 * 0.2 * 0.6) * 100 = 7.2$; $fm(MPshort) + fm(PPshort) = (0.25 * 0.2 * 0.6 + 0.2 * 0.2 * 0.6) * 100 = 5.4$, so $S(Pshort) = (40.2, 45.6]$;

$fm(LPshort) + fm(VLshort) = (0.2 * 0.2 * 0.6 + 0.35 * 0.2 * 0.6) * 100 = 6.6$; $fm(MLshort) + fm(PLshort) = (0.25 * 0.2 * 0.6 + 0.2 * 0.2 * 0.6) * 100 = 5.4$, so $S(Lshort) = (52.2, 57.6]$;

with similar calculations, we have $S(W) = (57.6, 61.6]$; $S(Llong) = (61.6, 65.2]$; $S(Plong) = (69.6, 73.2]$; $S(Mlong) = (78, 82.5]$; $S(Vlong) = (88.8, 95.1]$; $S(I) = (95.1, 100]$;

Step (21) - (28): Determine the *neighborhood level 2* of *Little Long* and *Little Short*. We have $Little Long \in S(Little Long)$ so *neighborhood level 2* of *Little Long* is $FRN_2(Little Long) = S(Little Long) = (61.6, 65.2]$, and *neighborhood level 2* of *Little Short* is $FRN_2(Little Short) = S(Little Short) = (52.2, 57.6]$.

Step (29) - (32): According to conditions:

- The length *Little Long* so we have two objects satisfied is iD1, iD5.
- The width *Little Short* so we have three objects satisfied is iD1, iD2, iD6.

So result = {iD1, iD2, iD5, iD6} satisfied a query with the operation *or*.

Query 2: List of rectangles have area is "*less small*". Using algorithms **SMSN** the following:

Step (1) - (6):

Let consider a linear hedge algebra of *length*, $X_{length} = (X_{length}, G_{length}, H_{length}, \leq)$, where $G_{length} = \{Short, Long\}$, $H_{length}^+ = \{More, Very\}$, $H_{length}^- = \{Possibly, Little\}$, where *P*, *L*, *M* and *V* stand for *Possibly*,

Little, *More* and *Very*, with *Very* > *More* and *Little* > *Possibly*.

Suppose that $W_{length} = 0.6$, $fm(short) = 0.6$, $fm(long) = 0.4$, $fm(V) = 0.35$, $fm(M) = 0.25$, $fm(P) = 0.2$, $fm(L) = 0.2$.

$Dom(DODAI) = [0, 100]$. $Result = \emptyset$;
 $LD_{length} = H_{length}(short) \cup H_{length}(long)$.

Step (7) - (20): so *less small* we see it corresponds to *Little Short*, that *Little Short* = 2 so we only need to build interval *similar level 2*. We perform partition the interval [0, 100] into interval *similar level 2*: (similar calculation in query 1)

$S(0) = [0, 7.35]$; $S(Vshort) = (7.35, 16.8]$; $S(Mshort) = (26.25, 33]$; $S(Pshort) = (40.2, 45.6]$; $S(Lshort) = (52.2, 57.6]$; $S(W) = (57.6, 61.6]$; $S(LLong) = (61.6, 65.2]$; $S(PLong) = (69.6, 73.2]$; $S(MLong) = (78, 82.5]$; $S(VLong) = (88.8, 95.1]$; $S(I) = (95.1, 100]$;

Step (21) - (32): Determine the *neighborhood level 2* of *less small*. So *less small* = *Little Short* $\in S(Little Short)$ so *neighborhood level 2* of *less small* is $FRN_2(Little Short) = S(Little Short) = (52.2, 57.6]$.

Step (33) - (40): According to conditions:

- The length *Little Short* so we have two objects satisfied is iD2, iD6.
- The width *Little Short* so we have three objects satisfied is iD1, iD2, iD6.

The function combined hedge algebra is product of hedge algebra with the operation *and*, so result = {iD2, iD6} satisfied the conditions of query 2.

V. CONCLUSION

In this paper, we propose a new method for linguistic data processing in object-oriented database that its information is fuzzy and uncertainty approach to the semantic neighborhood based on hedge algebras. This approach makes easy to process data and homogeneous data. Based on quantitative semantics, we determined *neighborhood level k* of linguistic values and perform data matching by *neighborhood level k* this. This paper has proposed a method combination of hedge algebras in case the attribute value is the linguistic value. From data matching based semantic neighborhood of hedge algebras, this paper has proposed two algorithms **SASN** and **SMSN** for searching data with fuzzy conditions based semantic neighborhood of hedge algebras.

REFERENCES

- [1]. Berzal, F., Martin N., Pons O., Vila M.A. A framework to build fuzzy object-oriented capabilities over an existing database system. In Ma, Z. (Ed): Advances in Fuzzy Object-Oriented Database: Modeling and Application. Ide Group Publishing, 2005a, 117-205.
- [2]. D.V.Thang, D.V.Ban. Query data with fuzzy information in object-oriented databases an approach

- interval values. International Journal of Computer Science and Information Security (IJCSIS), Vol.9.No2, 2011, pp 1-6.
- [3]. Le Tien Vuong, Ho Thuan, *A relational database extended by application of fuzzy set theory and linguistic variables*. Computer and Artificial Intelligence 8(2) (1989), pp 153-168.
- [4]. Ho Thuan, Ho Cam Ha, *An approach to extending the relational database model for handling incomplete information and data dependencies*. Journal of Computer Science & Cybernetic (3) (2001), pp 41-47.
- [5]. N.C. Ho, *Fuzzy set theory and soft computing technology. Fuzzy system, neural network and application*, Publishing science and technology 2001, p 37-74.
- [6]. N.C. Ho, *Quantifying Hedge Algebras and Interpolation Methods in Approximate Reasoning*, Proc. of the 5th Inter. Conf. on Fuzzy Information Processing, Beijing, March 1-4 (2003), p105-112.
- [7]. N. C. Ho, W.Wechler, "*Hedge Algebras: an algebraic approach to structure of sets of linguistic domains of linguistic truth variable*", Fuzzy Set and System, 35 (1990), pp 281-293.
- [8]. N.C.Ho, N.C.Hao, *A method of processing queries in fuzzy database approach to the semantic neighborhood of Hedge Algebras..* Journal of Computer Science & Cybernetic, T.24, S.4 (2008), pp 281-294.
- [9]. P.M.Tam, T.T.Son, *A fuzzy database and applications in criminal management*. Journal of Computer Science & Cybernetic, T.22, S.1 (2006), pp 67-73.

AUTHORS PROFILE

Name: Doan Van Thang

Birth date: 1976.

Graduation at Hue University of Sciences – Hue University, year 2000. Received a master's degree in 2005 at Hue University of Sciences – Hue University. Currently a PhD student at Institute of Information Technology, Academy Science and Technology of Viet Nam.

Research: Object-oriented database, fuzzy Object-oriented database. Hedge Algebras.

Email: vanthangdn@gmail.com

A Low-Power CMOS Programmable CNN Cell and its Application to Stability of CNN with Opposite-Sign Templates

S. El-Din, A. K. Abol Seoud, and A. El-Fahar
Electrical Engineering Department
University of Alexandria
Alexandria, Egypt.
E-mail: eng_salah_alx@yahoo.com

M. El-Sayed Ragab
School of Electronics, Comm. and Computer Eng.
E-JUST.
Alexandria, Egypt.
E-mail: m.ragab@ejust.edu.eg

Abstract--In this paper, a novel VLSI architecture adaptation of the Cellular Neural Network (CNN) paradigm is described. It is based on a combination of MOS transistors operating in weak inversion regime. This combination has enabled a CMOS implementation of a simplified version of the original CNN model with the main characteristics of low-power consumption. Digitally selectable template coefficients are employed and a local logic and memory are added into each cell providing a simple dual computing structure (analog and digital). A four-quadrant analog multiplier is used as a voltage controlled current source which is feeding from the weighting factors of the template elements. The main feature of the multiplier is the high value of the weight voltage range which varies between the ground voltage and the supply voltage. A simulation example for stability of a class of nonreciprocal cellular neural network with opposite-sign template is presented.

Keywords: Cellular Neural Network, Low-power CNN, Opposite-Sign Template.

I. INTRODUCTION

Cellular Neural Networks (CNNs), introduced by Chua and Yang in 1988 [1], have been extensively studied in the past two decade [2, 3, 4]. All such studies have been focused on four special topics: 1) the CNN functions; 2) hardware implementation; 3) software systems; and 4) various engineering and scientific applications [5]. CNNs have been successfully applied to signal processing systems, especially in static image treatment [3], and to solve nonlinear algebraic equations [6]. It has also been shown that the process of moving images requires the introduction of time delays in the signals transmitted through the network [7,8,9]. Through VLSI technology and using switching circuit techniques such delays can be introduce in the interaction between neurons [8]. To realize the CNN on a silicon chip, the CNN cell is required to have low power consumption. Various analog VLSI implementations of CNN building locks have been previously implemented and tested [10, 11]. Such implementations have served to build CNNs under different constraints concerning the size of the network, the kind of cell input and state (analog/digital), the power consumption, and the programmability features of the network allowing more compact VLSI implementations [12].

The aim of this paper is to design and implement a new low-power CMOS CNN cell. The circuit employs low-

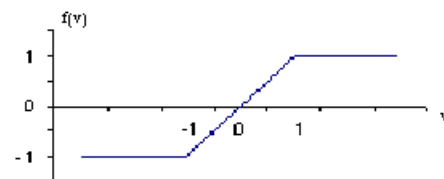
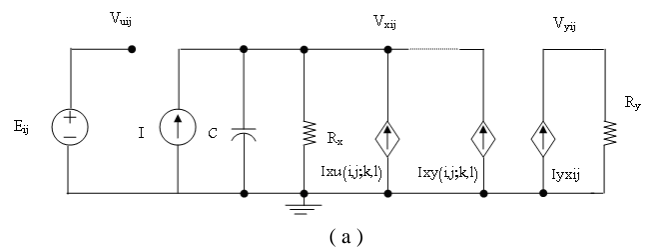
power four quadrant multipliers using MOSFET's operating in the weak inversion regime, where the small currents contribute to the low- power consumption [13]. The multiplier also has a variable transconductance characteristic for the programmability of the CNN structure. The proposed cell has been applied to study the stability [14, 15], and oscillation of a CNN paradigm [16, 17]. The performance of the proposed circuit has been evaluated using PSPICE simulations.

II. The General Framework

A cellular neural network [1] is a special type of neural networks, where the analog processing elements on one layer are arranged in a two-dimensional grid having cell interconnections with nearest neighbors only. Consider the analog processing cell circuit, henceforth called a cell, as shown in Fig.1(a), with only one nonlinear element whose characteristics is shown in Fig.1(b). This cell is located in the (i, j) position of a two-dimensional regular array of $M \times N$ cells. The r -neighborhood $N_r(i, j)$ of a typical cell $C(i, j)$ is defined as:

$$N_r(i, j) = \{ C(k, l), \max(|k-i|, |l-j|) = r \text{ (integer)} \} \quad (1)$$

An $r=1$ neighborhood of a cell within a cell array consists of all those cells shown shaded in Fig.1(c).



(b)

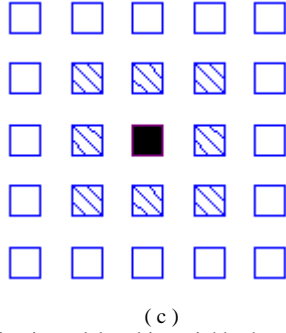


Figure 1. The cell circuit model and its neighborhood in a cell array. (a) The cell circuit model (b) The characteristics of the single nonlinear element of the cell (a voltage-controlled current source). (c) An $r=1$ neighborhood in a part of a cell array.

The dynamical system equations describing a cellular neural network consist of the following equations and constraints:

(1) State Equation:

$$C \frac{dV_{xij}}{dt} = -\frac{1}{R_x} V_{xij}(t) + \sum_{C(k,l) \in N_r(i,j)} A(i,j;k,l) V_{ykl}(t) + \sum_{C(k,l) \in N_r(i,j)} B(i,j;k,l) V_{ukl}(t) + I \quad (2)$$

where

$$1 \leq i \leq M; \quad 1 \leq j \leq N$$

(2) Output Equation:

$$V_{yij}(t) = 0.5 \left(|V_{xij}(t) + 1| - |V_{xij}(t) - 1| \right) = f(V_{xij}). \quad (3)$$

(3) Input Equation:

$$V_{uij} = E_{ij}. \quad (4)$$

(4) Constraint Equations:

$$|V_{xij}(0)| \leq 1 \quad (5)$$

$$|V_{uij}| \leq 1. \quad (6)$$

(5) Parameter Assumptions:

$$A(i,j;k,l) = A(k,l;i,j) \quad \text{Symmetry condition} \quad (7)$$

$$C > 0, \quad R_x > 0. \quad (8)$$

III. Stability of Cellular Neural Networks

A necessary condition for the proper operation of a cellular neural network is that it be completely stable within the dynamic range of prescribed inputs. A circuit is said to be completely stable if every trajectory tends to an equilibrium state. The complete stability of a subclass of cellular neural networks is defined by symmetric templates [1]. The symmetry condition means that the feedback values between any two cells are reciprocal in the sense that corresponding values are the same; i.e., $A(i,j;k,l) = A(k,l;i,j)$. The assumption (7) implies the perfect symmetry of the feedback-template values between any two cells within a neighborhood. From theorem 4 in [1], if the parameters satisfy the symmetry condition, the circuit will be completely stable. But many unsymmetrical templates have been found for some important applications [3]. In [14] it has been shown that for a class of practically important templates (positive / negative and opposite-sign templates), the complete stability property

is assured even if the symmetry condition is not met. In [15] a through stability analysis of cellular neural networks with opposite-sign templates has been presented. In this analysis, the dependence of complete stability on the template values, and the parameter regions for complete stability and instability have been determined. This class is defined by the template values which satisfy the following structures and sign conditions [14]:

$$A = \begin{bmatrix} 0 & 0 & 0 \\ s & p & -s \\ 0 & 0 & 0 \end{bmatrix} \quad (9)$$

$$\text{where } p > \frac{1}{R_x} \text{ and } s > 0$$

The complete stability of the system defined by (2) has been proven to be strongly conjectural if [15]:

$$i) \quad B=0 \quad ii) \quad \frac{(p-1)}{2} \leq s \leq (p-1) \quad (10)$$

Also, the network will oscillate periodically if [16]:

$$i) \quad B=0 \quad ii) \quad s > p-1 \quad (11)$$

IV. Low-Power CMOS Programmable CNN Cell

The block diagram of a continuous time CNN cell is shown in Fig.2.

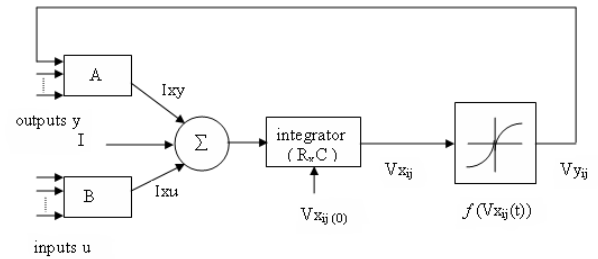


Figure2. Block diagram of CNN cell.

V_{xij} is the state of cell C_{ij} , with an initial condition $V_{xij}(0)$, $R_x C$ conforms the integration time constant of the system. The cell output is $V_{yij}(t) = f(V_{xij}(t))$, where f can be any convenient non-linear function. The block A can be implemented using a set of four quadrant multipliers whose inputs are the outputs of the cells within the assumed neighborhood and the template A values. Similarly, block B can be implemented using a set of four quadrant multipliers whose inputs are the inputs of the cells within the assumed neighborhood and the template B values. The outputs of blocks A and B are (in the current form) I_{xy} and I_{xu} , respectively. Those currents are summed with the bias current I of the cell and then integrated in the $R_x C$ circuit, to result in the cell state voltage V_{xij} . The output voltage of the cell V_{yij} is obtained through the limiting transfer function $f(V_{xij})$. Alternatively, the nonlinear transfer function $f(V_{xij})$ can be incorporated in the multiplier circuits themselves, resulting in a small area CNN cell. This can be realized using low-power CMOS four quadrant multipliers operating in weak inversion regime.

A. Programmable Low-power CMOS four quadrant multiplier

Fig. 3 shows the proposed programmable low-power CMOS four quadrant multiplier circuit and its sub-circuit representation.

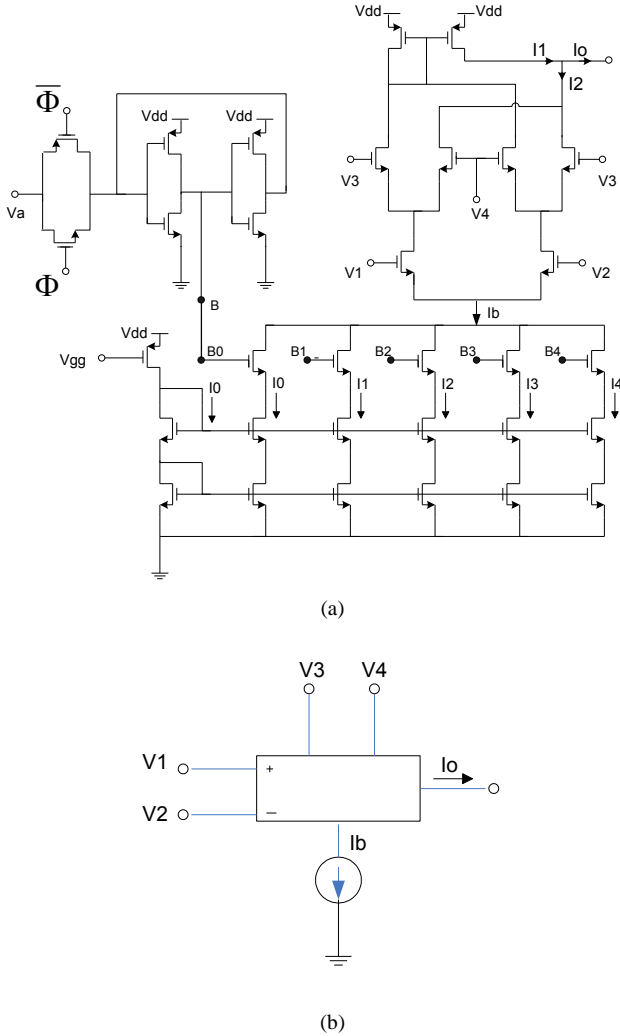


Figure 3. (a) Four quadrant multiplier schematic (b) Multiplier sub-circuit representation

This circuit represents a trade-off between digital and analog techniques. It is composed of registers which store the weight values, a linear DAC and a transconductance multiplier. The DAC has five bits plus sign weight storage which sets the tail bias current I_b . The least significant bit bias current has been set to 40 pA. The DAC has shown good monotonicity in the weak inversion regime. Each bit (B0-B4) of the DAC is controlled by a pass transistor which can be turned on or off depending on the value stored in the corresponding CMOS latch. I_0 - I_4 are the current sources which contribute to the bias current I_b in a successive power of two fashion. The DAC is connected to a transconductance amplifier to form a four quadrant multiplier. Assuming weak inversion operation for all MOS devices in the multiplier circuit, it can be shown that the output current I_o is expressed as:

$$I_o = I_1 - I_2 = \begin{cases} +I_b \tanh(k \frac{V_1 - V_2}{2}) & \text{if } V_3 \text{ is high and } V_4 \text{ is low} \\ -I_b \tanh(k \frac{V_1 - V_2}{2}) & \text{if } V_3 \text{ is low and } V_4 \text{ is high} \end{cases} \quad (12)$$

where $k = \frac{1}{nU_T}$, with n is a slope factor (in practice it

lies between 1 and 2 and is close to 1 for high values of gate voltage), and U_T is the thermal voltage whose value is 26mV at room temperature. Current switching logic controlled by V_3 and V_4 enables the output to change sign. The transfer characteristic of the multiplier circuit is shown in figure 4. It is noted that the output transfer characteristic is linearly proportional to one of the multiplier inputs, I_b , and varies nonlinearly with the other input, $(V_1 - V_2)$.

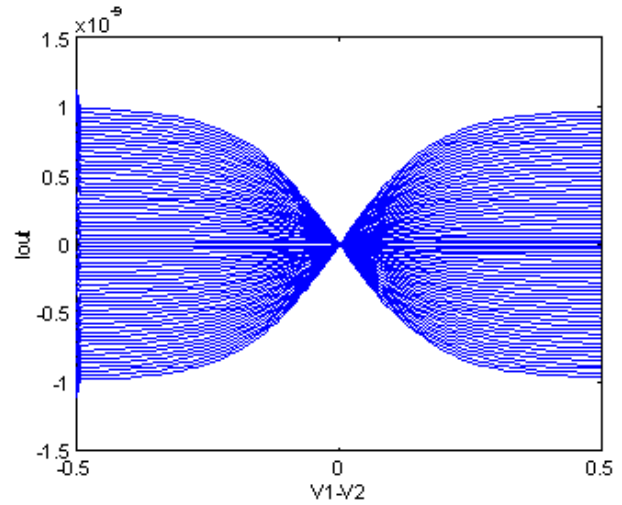


Figure 4. Transfer characteristic of the proposed four quadrant multiplier.

B. Complete CNN CMOS Implementation

Fig. 5 shows a complete implementation of a CNN cell using the proposed multiplier circuit.

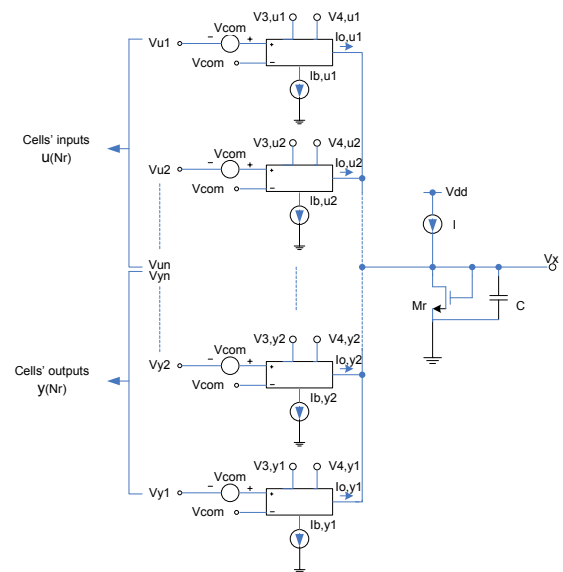


Figure 5. Complete CNN cell.

The sets of multipliers in the lower and upper parts of Fig.5 represent the second and third terms in the left hand side of equation (2), respectively. Each multiplier in the lower set accepts one of the cells' outputs within the given neighborhood, as one input, and the corresponding template value $A()$ as the other input. The A - template values are determined by the programmable tail current sources $I_{b,y}$ and their signs are controlled by the multiplier control inputs V_3 's and V_4 's. On the other hand, each multiplier in the upper set accepts one of the cell's inputs within the given neighborhood as one input, and the corresponding template value $B()$ as the other input. Also, those B - template values are determined by the programmable tail current source, $I_{b,u}$ and their signs are controlled by the corresponding multiplier control inputs V_3 's and V_4 's. The output currents of the two multiplier sets are summed together and applied to the $R_x C$ current integrator. The resistor R_x is implemented using the diode-connected transistor M_r .

V. SIMULATION EXAMPLE

To test the validity of the proposed CNN cell, a cellular neural network with two cells using an opposite-sign template is considered [15]. The network is shown in Fig.6.

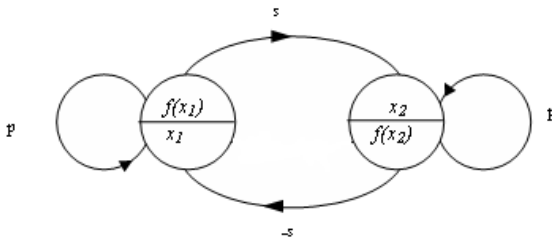


Figure 6. CNN with opposite-sign template.

The cells' inputs and their bias terms are set to zero, in order to remove the possibility of forced stability. The state equations of the system can be described by:

$$\begin{aligned} \dot{x}_1 + x_1 &= p f(x_1) - s f(x_2) \\ \dot{x}_2 + x_2 &= s f(x_1) + p f(x_2). \end{aligned} \quad (13)$$

Fig. 7 shows how the CNN of Fig.6 can be implemented using the proposed CNN cell. Note that in such an architecture the cell's state voltages V_{x1} and V_{x2} are directly fed back to the two cells and the nonlinear functions $f(x_1)$ and $f(x_2)$ are already embedded in the multipliers' transfer characteristics. As previously stated, this would guarantee compact CNN design architectures. The state equations resulting from such an implementation are then expressed as:

$$C \frac{dV_{x1}}{dt} = -\frac{1}{R_x} V_{x1} + a_1 I_{b0} f(V_{x1}) - a_2 I_{b0} f(V_{x2}) \quad (14)$$

$$\text{and, } C \frac{dV_{x2}}{dt} = -\frac{1}{R_x} V_{x2} + a_1 I_{b0} f(V_{x2}) + a_2 I_{b0} f(V_{x1}) \quad (15)$$

where R_x is the resistance of the diode-connected transistor,

$$f(V_x) = \tanh\left(\frac{kV_x}{2}\right),$$

and a_1 and a_2 represent the template-A values of the network.

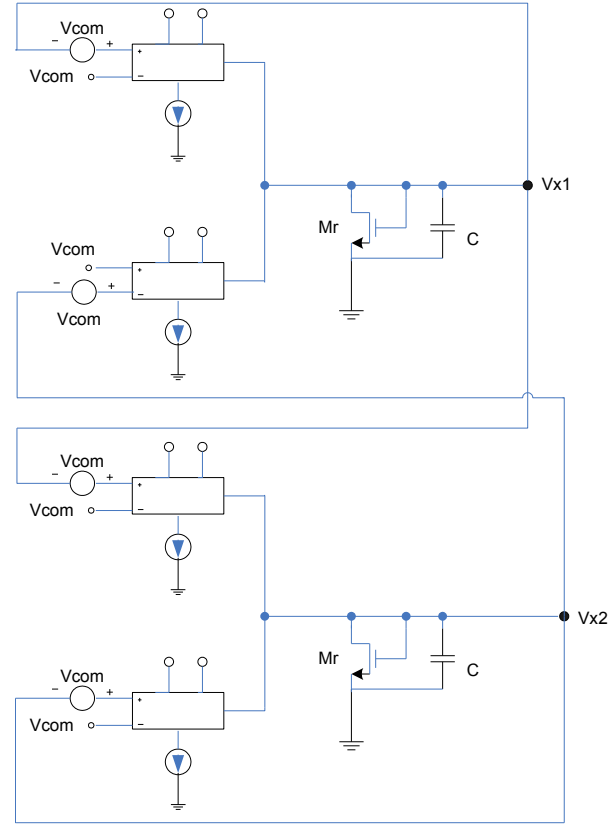


Figure7. Low-power CMOS implementation of the CNN of Fig. 6

AS mentioned previously, the network is considered to be

conjecturally stable if $\frac{(a_1-1)}{2} \leq a_2 \leq (a_1-1)$ and will

oscillate periodically if $a_2 > (a_1-1) > 0$. Fig .8 shows the transient behavior of the network with $a_1=2.0$ and $a_2=0.99$. A complete stability is observed in such a case where state voltages V_{x1} and V_{x2} are converged to constant values. Fig. 9 Shows the transient behavior of the network with $a_1=2.0$ and $a_2=1.2$. Periodic oscillations of state voltages V_{x1} and V_{x2} are observed in this case.

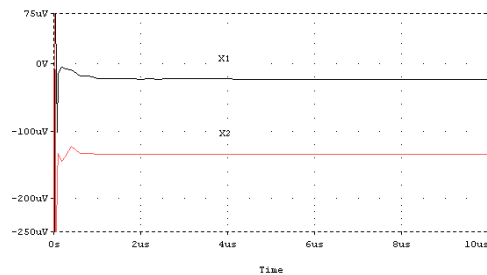


Figure8. Transient behavior of the network with $a_1=2.0, a_2=0.99$.

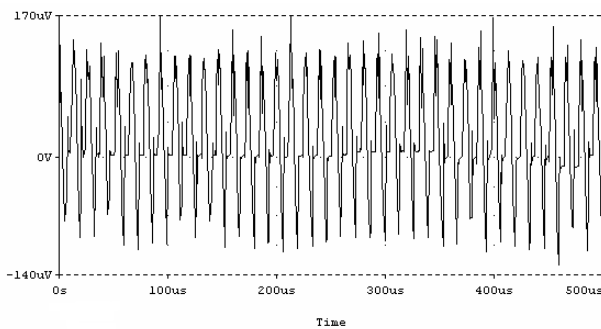


Figure 9. Transient behavior of the network with $a_1=2$, $a_2=1.2$.

VI. Conclusion

A modified low-power CMOS implementation of a cellular neural network cell has been proposed. Instead of the conventional piecewise-linear transfer function used in the output stage of the standard CNN cell introduced by Chua and Yang, a sigmoid-like transfer function is embedded in the transfer characteristic of the dependent current sources determining the state of the cell. This has been achieved by implementing those current sources using four-quadrant multipliers employing MOSFETs operating in weak inversion regime. The proposed CNN cell has been used to implement a complete CNN to study stability of a class of nonreciprocal CNN with opposite-sign template. The results have been confirmed using PSPICE simulations.

REFERENCES

- [1] L.O.Chua and L.Yang, "Cellular Neural Network: Theory," IEEE Trans. Circuits & Syst., Vol. 35, no. 10, pp. 1257-1272, oct.1988.
- [2] L.O.Chua and L.Yang, "Cellular Neural Networks: Applications," IEEE Trans. Circuits & Syst., Vol. 35, no. 10, pp 1273-1290, oct. 1988.
- [3] L.O.Chua, CNN: A paradigm for Complexity. Singapore: World Scientific, 1988.
- [4] J.M.Cruz and L.O.Chua, "Design of high-speed, high density CNNs in CMOS technology," in Cellular Neural Networks, T.Roska and J.Vandewalle, Eds. New York: Wiley, 1993, pp. 117-134.
- [5] X.Y.Wu, H.Chen, W.Tran, and W.K.Chen, "The improvement of the hardware design of artificial CNN and a fast algorithm of CNN component detector," J.Franklin Inst., Vol.330, pp.1005-1015, 1993.
- [6] L.O.Chua and L.Yang, "Equilibrium analysis of delayed CNNs," IEEE Trans. Circuits Syst., Vol.45, pp. 168-171, Feb. 1998.
- [7] P.P.Civalleri, L.M.Gilli, and L.Padolfi, "On stability of cellular neural networks with delay," IEEE Trans. Circuits Syst., Vol. 40, pp. 157-164, Mar. 1993.
- [8] S.Arik and V.Tavanoglu, "Equilibrium analysis of non symmetric CNN's," Int. J. Circuit Theory Applicat., Vol. 34, pp. 269-274, 1996.
- [9] Teh-lu Liao and Fong-Chin Wang "Global stability for Cellular Neural Networks with Time Delay" IEEE Transactions on neural networks, Vol. 11, no. 6, November 2000.
- [10] J.M.Cruz and L.O.Chua, "A CNN chip for connected component detection," IEEE Trans. Circuits Syst., Vol. 38, pp. 812-816, July 1991.
- [11] P.Kinget and M.S.J.Steyaert, "A programmable analog cellular neural network CMOS chip for high speed image processing," IEEE. J. Solid-State Circuits, Vol. 30, Mar. 1995.
- [12] Mancina Anguita, Francisco. J.Pelayo, Francisco J.Fernandez, and Alberto Prieto "A Low-Power CMOS Implementation of Programmable CNN's with Embedded Photo sensors" IEEE Transactions on circuits and systems: Fundamental theory and applications; Vol. 44, no. 2, Feb. 1997.
- [13] Salah el-Din "Design and Simulation of Compact Low-Power CMOS Artificial Neural Networks" M.SC. dissertation, Alexandria University, Alex., Egypt, 2003.
- [14] L.O.Chua, Fellow, IEEE, and Tamas Roska, Member,IEEE, "Stability of a Class of Nonreciprocal Cellular Neural Networks," IEEE Trans. Circuits & Syst. vol., 37, no. 12, December 1990.
- [15] Fan Zou and Josef A.Nossek "Stability of Cellular Neural Networks with Opposite-Sign Templates," IEEE Trans. Circuits & Syst. vol. 38, no. 6, June 1991.
- [16] Fan Zou and Josef A.Nossek, "A Chaotic Attractor with Cellular Neural Networks," IEEE Trans. Circuits& Syst., vol.38, no.7, pp.811-812, July 1991.
- [17] Hisashi Tanaka, Koichi Tanno, Hiroki Tamura and Kenji Murao "Low-Power CMOS CNN Cell and its Application to an Oscillatory CNN," The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008).
- [18] M.A.Jabri,R.J.Cggins and B.G.Flower "Adaptive Analog VLSI Neural Systems" SEDAL, CHAPMAN & HALL -1996.

A Novel Model for Synchronization and Positioning by using Neural Networks

Hossein Ghayoumi Zadeh*, Siamak Janianpour and Javad Haddadnia

Department of Electrical Engineering
Sabzevar Tarbiat Moallem University
Sabzevar, Khorasan Razavi, Iran
Email: h.ghayoumizadeh@gmail.com*

Abstract— In this paper by using a Low Noise Amplifier (LNA), a synchronization and positioning system is designed. Parameters that cause the system to be far from ideal condition such as S-Parameters, Noise Figure, IIP3, and Gain are considered that is one of the advantages of this system. In this stage this process is a little slow so by adding the neural network to the system the speed of synchronization is increased. By using the neural network the time needed to calculate the time difference of arrival (TDOA) is significantly decreases.

Keywords—component; Global Positioning System, Low-noise amplifiers, Neural networks.

I. INTRODUCTION

The TDOA approximation has so many different applications such as communication, electronic war and medical engineering. Following some of these applications will be discussed. One of the most important applications of TDOA is in positioning of the transmitters. Nowadays radar and sonar systems are widely used with many different military or nonmilitary applications and their importance in security problems are so that they are parts of the strategic system of each country so to protect the radars; the usage of the passive radar is become popular, increasingly[1]. Although in these type of radars the basic principle of the radars are dominant but the transmitter of the radar is omitted from the system and by omitting the transmitter the receiver will become hidden from the sight of the enemies. The place of transmitter is one of the most important parameters that assign the duties of the radar. Another application of TDOA is in measurement and controlling the coolant current of the atomic reactors. Also it can be used to localize the position of brain that controls the simultaneity of the activities in epileptic patients. Further it is

used in global positioning system (GPS), recently. Contrary to other methods, this system will not affect the normal operation of the satellites, because the time delay is calculated passively. Also there is no need to carry extra hardware in spaceships and this will reduce the cost of this procedure.

By calculating the received uplink signal TDOA using three or four satellite on earth orbit, the position of the transmitters can be localized. When the transmitter is located on the earth, three satellites and when the elevation of the transmitter is not known, four satellites are needed to localize the position of the transmitters. So in the passive systems, approximation of the delay time between the signals receive time from two different sensors plays a significant role in measurement of the distance and direction of the transmitters. When a signal is emitted into the environment from transmitter, it spreads with a specific speed, so two receivers with different distance from the transmitter, sense the signal with a time delay. If $S(t)$ is the emitted signal from transmitter and assume that there is just one way for signal transmission, then the signals first will receive to the nearest receiver and with a delay to the next receiver, the delay time is shown with D . So the goal is to measure this delay and approximation of this delay has a significant role in synchronization and positioning process [2].

II. SYNCHRONIZATION

The controll system is shown in fig. 1. In this circuit the LNA model is used in S-parameters block. In this block the values of S-parameters, Noise Figure and IIP3 are changeable.

In this circuit the Gaussian Signal is used as input (Fig. 2). Also the white noise is applied to the signal and considered as a non ideal factor.

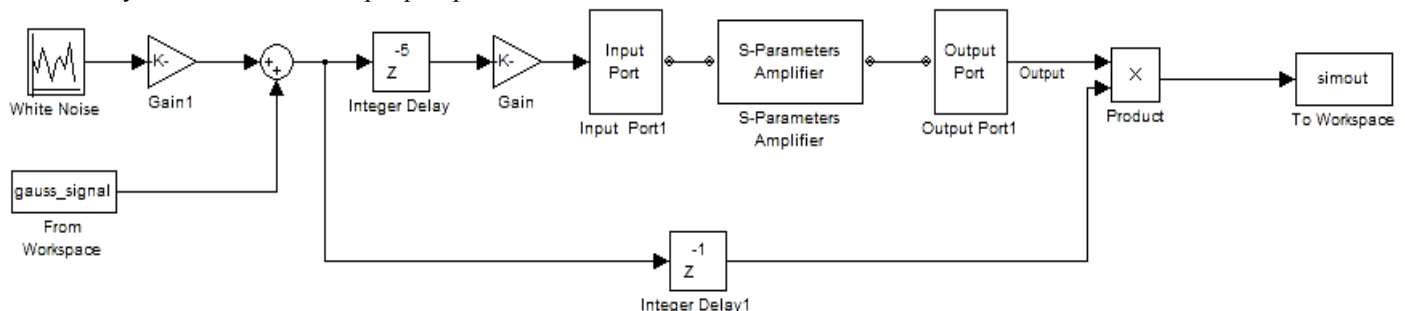


Figure 1. Controlling system designed for Synchronization

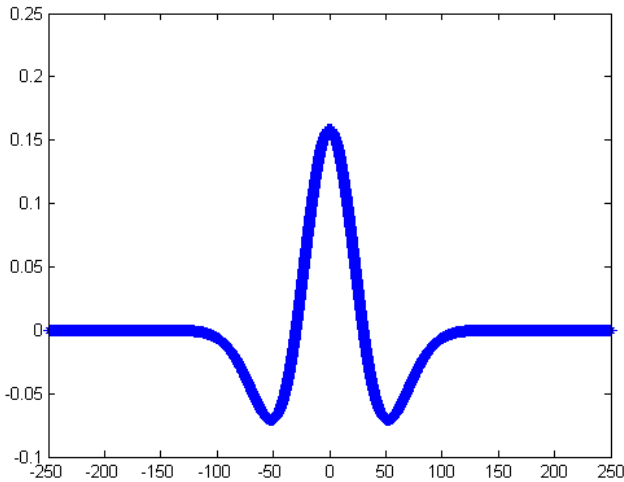


Figure 2. The Gaussian Signal that is applied to the control circuit as input

The structure of the synchronization is so that the input delay of T1 is considered by the integer delay block. After the amplification stage, the changes of the S-parameter are exerted to the input signal that simulates the LNA stage, and it is transferred to the output. The second input in this simulation that is also the Gaussian signal, is multiply to the first signal with a delay and it is transferred to the output. The point is that the second output also has delay equal to T2 that is shown in the “integer delay 2” block. Now by using a feedback from the output to the second input, the delay is changeable so that the both inputs become concurrent.

The easiest and most effective way for synchronization of the both input signals is to multiply the feedback output with the first input, and simultaneously check the output until the output becomes maximal. Now the delay can be recorded and stored. When the both inputs become synchronize the output will become maximal.

III. LOW NOISE AMPLIFIER (LNA)

In this paper the model of an inductorless low-noise amplifier (LNA) is used. [3] This LNA is designed for ultra-wideband (UWB) receivers and for microwave access, covering the frequency range from 0.4 to 5.7 GHz using 0.18- μ m CMOS technology. Simulation results show that the voltage gain reaches a peak of 19.6 dB in-band with an upper

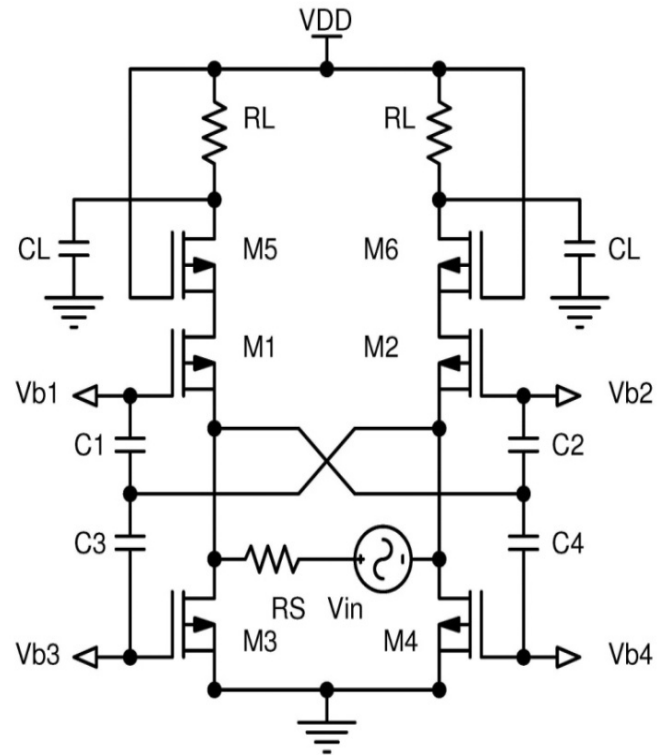


Figure 3. LNA circuit that is used in the controlling circuit.

3-dB frequency of 5.7 GHz. The IIP3 is about -3 dBm and the noise figure (NF) ranges from 3.06-3.8 dB over the band of interest. Input reflection coefficient S11 is below -8.79dB for the design. The LNA consumes 5.77 mW from a low supply voltage of 1.8 V. A figure of merit is devised to compare the proposed designs to recently published wideband CMOS LNAs. The proposed topology achieves a lower NF than that of the topology capacitive cross-coupling with inductors, with the additional advantage of removing the bulky inductors. It is shown that the LNA is designed without on-chip inductors that its performance is comparable with inductor-based designs. The LNA circuit that is used in this system is shown in fig. 3.

LNA properties such as gain, IIP3 and etc. that are used for synchronization in this system are given in TABLE I.

Also an example of the test results is shown in TABLE II.

TABLE I. LNA PROPERTIES THAT ARE USED IN SIMULATION

Technology	BW(GHz)	NF (dB)	S ₁₁ (dB)	Gain (db)	IIP3 (Dbm)	Power(mW)	No.of Coils	FOM
0.18 μ m	0.4 ~ 5.7	3.06 ~ 3.8	<-8.79	19.66	-3	5.77	0	3.68

TABLE II. TEST RESULTS

Output	Gain1	Delay	Gain	Noise Figure (db)	OIP3(dbm)	Frequency	S11	S12	S21	S22
10	-10 ⁻⁸	5	100	0	inf	3.5E6	-9.32	1.76	-49.76	-8.64

By evaluation of the test results, some points are achievable:

1. The delay time is respectively related to noise, noise figure, initial delay and OIP3. The LNA gain will affect the delay time but it has a little impact. Also the impact of S-parameter is so small that it is neglect able.
2. By increasing the noise figure the delay time will increase. Also in high value of the LNA gain its increase will cause an increase in delay time, but noise figure does not show such a behavior.
3. The effect of the OIP3 in comparison with S-parameter is higher. (by changing the S-parameter values not a big change is observed).
4. By increasing the OIP3 value the delay time is decreased.

But the problem in designing such a controlling system is that about 45 second is needed to process such a big amount of data and it is one of the disadvantages of this system, because moreover to input delay time we will lose the time needed for the calculation process of software that is very bad for synchronization systems. So to reduce the calculation time the neural network is used.

IV. USAGE OF NEURAL NETWORK IN SYNCHRONIZATION

In this system a linear neural network is used. To minimize the errors in these networks, the training process is done by using squares least mean algorithm (Fig. 4).

The neural network that is used in this process must be so fast and so accurate. First of all the goal is to synchronize the signals with each other by using the neural network. To train the neural network the previous Gaussian signal is used that is shown in fig. 5.

The objective function of the neural network can be achieved from multiplying of two Gaussian signals that is shown in fig. 6.

Now many samples of Gaussian signal with different delay time are applied to the neural network for training purpose. For instance to test the designed neural network the signal that is shown in fig. 7 is used. Also to test the robustness of the network, samples that contain noise and have different delays in comparison with original signals are used.

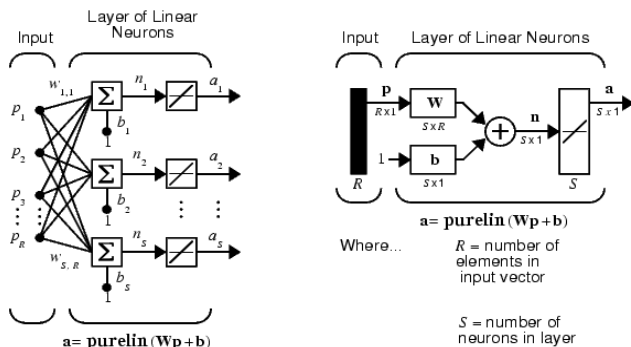


Figure 4. The structure of linear neural network used in this control system.

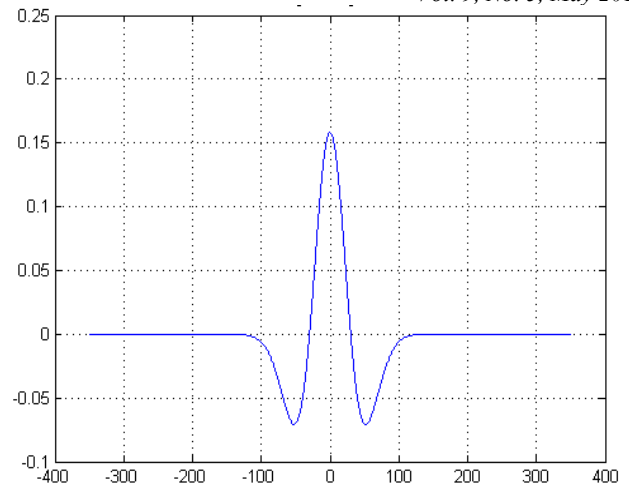


Figure 5. The initial signal that is used to train the neural network

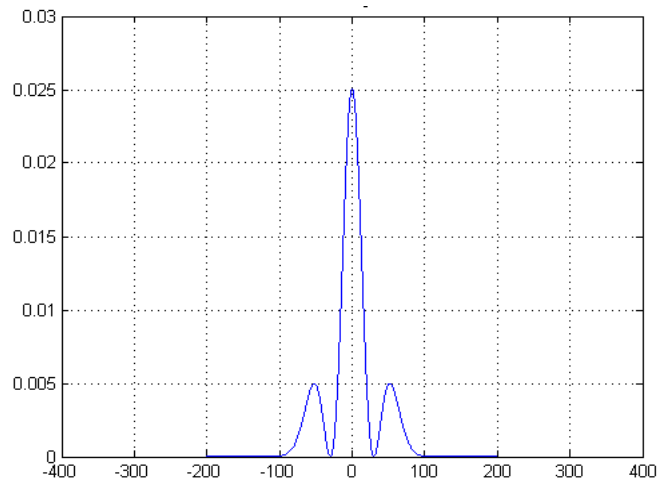


Figure 6. The objective signal that is used to train the system.

An example of the signal that is used for test purpose is shown in fig. 7. The noises and delays are obviously shown in this figure.

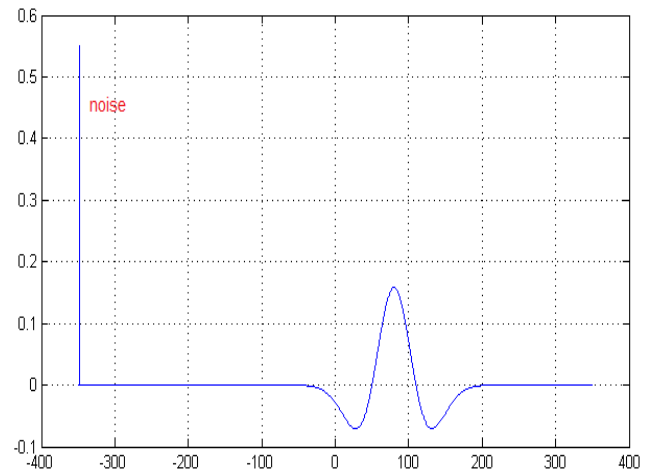


Figure 7. The test signal that contains noise and delay

Now the results achieved by using neural networks are shown in fig. 8. The neural network tries to find the appropriate delay time that leads to the objective function by using the iterative methods. If you look at the fig. 8 carefully you will see that some parts of the chart are drawn thicker and or bolded. That is caused by increasing the accuracy of the neural network and the number of iterations. In other words, in this method two signals are studied by different random delays and the delay that cause the output to reach the objective function is recorded.

In the proposed method the noise in the Gaussian signal, has a small effect on calculation of the delay time that is clearly shown in fig. 8.

By decreasing the accuracy of the network the segregation of the signals with different delay time are shown more clearly (fig. 9).

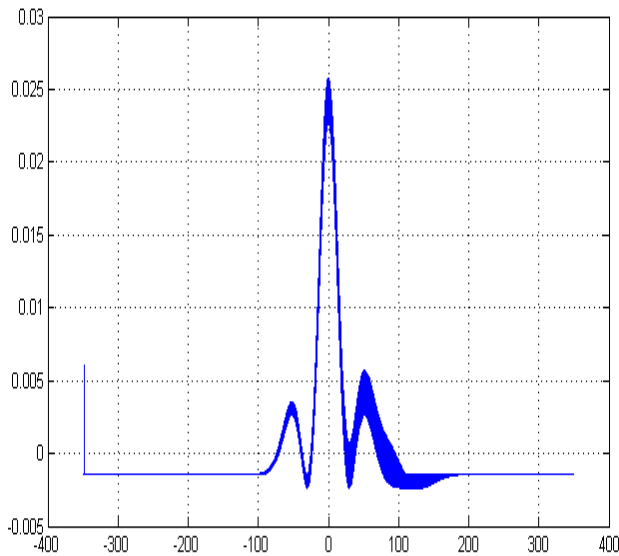


Figure 8. The output of the neural network.

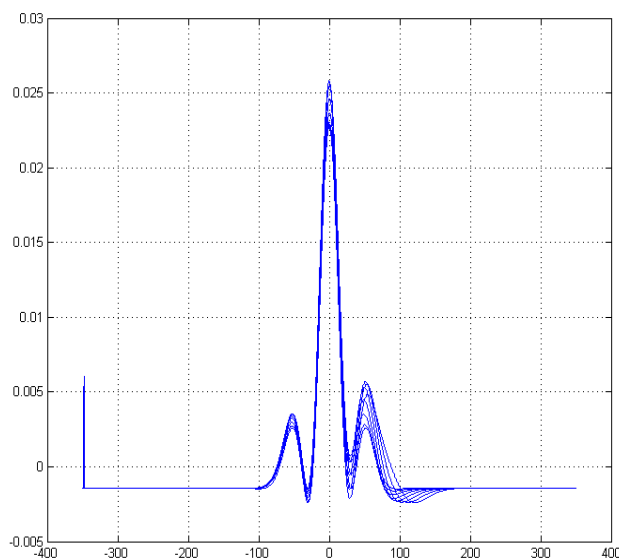


Figure 9. The more detailed chart of the output signal.

In this method the speed of delay calculation is significantly increased and the calculation time is decreased. It is good to mention that the Gaussian signal enters the system periodically and the system must be able to sweep the input continuously. To consider the noise in this system, for synchronization other techniques must be added to this system. Now we are going to describe these techniques. Assume a signal like the signal in fig. 10 is given to the system as input. First an intact period of the signal must be given to the system as the training sample.

V. USAGE OF NEURAL NETWORK IN POSITIONING

There are so many methods for positioning systems based on the calculation of the signal time-of-flight. One of these methods is TDOA. In this method to calculate the position of the TAG, the distance between the TAG and a node is calculated and it is compared with the distance of the TAG with another node. In this method the TAGs are just the transmitter and the nodes are just receiver. In 2 dimensional systems 4 TAGs and in 3 dimensional systems 5 TAGs are used.

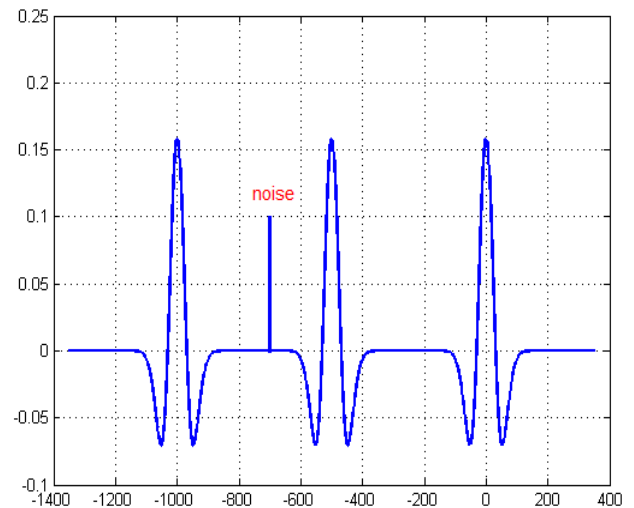


Figure 10. Three period of signal with consideration the effect of noise.

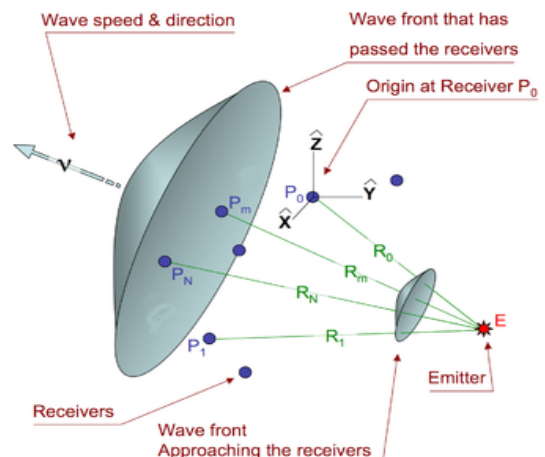


Figure 11. Geometrical structure of a 3D TDOA.

In this study the unknown position of the TAG is shown with function E (Equation 1).

$$E=(x,y,z) \quad (1)$$

And the receivers are defined as (Equation 2):

$$P_0, P_1, \dots, P_m, \dots, P_N, P_m = (x_m, y_m, z_m), 0 \leq m \leq N \quad (2)$$

Where N is the function dimension.

The distance between the each transmitter and the receiver is defined as R_m and R_0 is the distance between transmitter and the false origin (it is assumed that one of the receivers is located at the false origin). As it is shown in fig. 11 the resultant time is calculated as follows (Equation 3):

$$\mathcal{G}\tau_m = \mathcal{G}T_m - \mathcal{G}T_0 \quad (3)$$

In this equation, $\mathcal{G}\tau_m$ is the time that the signal needs to arrive to the m^{th} receiver. The delay time that is considered in this stage must refer to the calculated time in synchronization stage. Now, the time duration (τ_m) can be calculated by using correlation function ($P_m \times P_0$).

Substituting R_0 instead of $\mathcal{G}T_0$ in above equation and after some rearrangements, finally we will have a line equation with constant coefficients (equation 4)[4,5].

$$A_m x + B_m y + C_m z + D_m = 0 \quad (4)$$

That A_m , B_m , C_m and D_m in equation 4 are defined as follows.

$$A_m = \frac{2x_m}{\mathcal{G}\tau_m} - \frac{2x_1}{\mathcal{G}\tau_1} \quad (4a)$$

$$B_m = \frac{2y_m}{\mathcal{G}\tau_m} - \frac{2y_1}{\mathcal{G}\tau_1} \quad (4b)$$

$$C_m = \frac{2z_m}{\mathcal{G}\tau_m} - \frac{2z_1}{\mathcal{G}\tau_1} \quad (4c)$$

$$D_m = \mathcal{G}\tau_m - \mathcal{G}\tau_1 - \frac{x_m^2 + y_m^2 + z_m^2}{\mathcal{G}\tau_m} + \frac{x_1^2 + y_1^2 + z_1^2}{\mathcal{G}\tau_1} \quad (4d)$$

It is very easy to calculate the three unknown parameters (x,y,z) that are the coordinates position of the transmitter, by solving the three equations with one of the different methods such as singular value decomposition, numerical analysis or

etc. In this study the numerical analysis is used. TDOA is mostly used in open areas for personal uses, air traffic control (ATC) or military systems.

A sample of positioning process is shown in fig. 12.

Because of linearity of the equations, theoretically there isn't any error in this process but the practical problem, is the signal transfer time calculation from the transmitter to the receiver. In this paper the role of clock pulse generation and synchronization is significant. To increase the accuracy of the system, the clock pulse must be synchronized in every node. If the generated clock pulse was in the order of 1ns, the best achievable accuracy in the positioning system is about 30 centimeter that means we can find the position of the transmitter with maximum error of 30 centimeter.

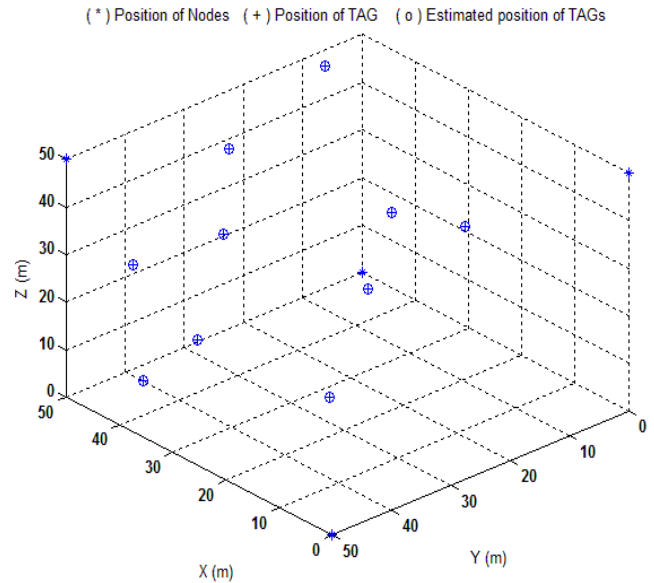


Figure 12. Positioning with TDOA method

REFERENCES

- [1] Fujiwara, R.; Mizugaki, K.; Nakagawa, T.; Maeda, D.; Miyazaki, M.; , "TOA/TDOA hybrid relative positioning system using UWB-IR," Radio and Wireless Symposium, 2009. RWS '09. IEEE , pp.679-682, 18-22 Jan. 2009.
- [2] Dawei, L.; , "Application of assisted TDOA positioning technology in vehicle positioning and navigation," Mobile Technology, Applications and Systems, 2005 2nd International Conference on , pp.5 pp.-5, 15-17 Nov. 2005.
- [3] Ali Shirzad Nilsaz, Mohsen Khani Parashkoh, Hossain ghauomy-zadeh, Zhuo, Zou, Majid Baghaei-Nejad, Li-Rong Zheng. "low power .18um cmos ultra wideband inductor-less lna design for uwb receiver".Asia Pacific Conference on Circuits and Systems,2010.
- [4] Kian Meng Tan; Choi Look Law; , "GPS and UWB Integration for indoor positioning," Information, Communications & Signal Processing, 2007 6th International Conference on , pp.1-5, 10-13 Dec. 2007.
- [5] Rabinowitz, M.; Spilker, J.J., Jr.; , "A new positioning system using television synchronization signals," Broadcasting, IEEE Transactions on , vol.51, no.1, pp. 51- 61, March 2005.
- [6] Wenjuan Jia; Chunlan Yang; Guocheng Zhong; Mengying Zhou; Shuicai Wu; , "Fetal ECG extraction based on adaptive linear neural network," Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on , vol.2, pp.899-902, 16-18 Oct. 2010.

AUTHORS PROFILE



Hossein Ghayoumi Zadeh received the B.Sc. degree in electrical engineering with honors from Sabzevar Tarbiat Moallem University, Sabzevar, Iran, in 2008. He is now M.Sc. student in electrical and electronic engineering at Sabzevar Tarbiat Moallem University in Iran. His current research interests include computer vision, pattern recognition, image processing, artificial neural network, intelligent systems, fuzzy logic and soft computing and etc.



Javad Haddadnia received his B.Sc. and M.Sc. degrees in electrical and electronic engineering with the first rank from Amirkabir University of Technology, Tehran, Iran, in 1993 and 1995, respectively. He received his Ph.D. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran in 2002. He joined Tarbiat Moallem University of Sabzevar in Iran since 2002 as an associated professor. His research interests include neural network, digital image processing, computer vision and medical Engineering. He has published several papers in these areas. He has served as a Visiting Research Scholar at the University of Windsor, Canada during 2001- 2002. He is a member of SPIE, CIPPR, and IEICE.



Siamak Janianpour received the B.Sc. degree in mechanical engineering with honors from the Islamic Azad University Tehran south branch, Tehran, Iran, in 2008. He is now M.Sc. student in electrical and electronic engineering at Sabzevar Tarbiat Moallem University in Iran. His current research interests are computer vision, pattern recognition, digital image processing and analysis, intelligent systems, intelligent healthcare systems and etc. He is a member of the IEEE.

Strategic Approach for Automatic Text Summarization

Mr. Ramesh Vaishya

Sr. Lecturer, Department of Computer Science & Engg
Babu Banarsi Das National Institute of Technology &
Management
Lucknow, India
bbdnitm.rv@gmail.com

Dr. Surya Prakash Tripathi

Associate Professor, Department of Computer Science &
Engg
Institute of Engineering Technology
Lucknow, India
tripathee_sp@yahoo.co.in

Abstract— As the amount of information is increasing all the time, information modeling and analysis have become essential areas in information management. Information retrieval and storage is an essential part of Information processing. The major part of our useful information is in the form of text. Textual data which an individual goes through during daily processing are quite bulky and voluminous. The user can find the document from their internet and analyze all to sort out the relevant information. Analyzing the text by reading all textual data is infeasible. So the technology of automatic document summarizer may provide a solution to information overload problems. We propose an extractive text summarization system. Extractive summarization works by selecting a subset of sentences from the original text. Thus the system needs to identify most important sentences in the text. In our proposed work is to finding the important sentences using statistical properties like frequency of word, occurrence of important information in the form of numerical data, proper noun, keyword and sentence similarity factor. It depends on the net information content a particular sentence has. Any sentence having higher value is more relevance with respect to summary. Sentences are then selected for inclusion in the summary depending upon their relative importance in the conceptual network. The sentences (nodes in graph) are then selected for inclusion in final summary based on relative importance of sentence in the graph and weighted sum of attached feature score.

I. INTRODUCTION

We are drowning in information but starving for knowledge. Information is only useful when it can be located and synthesized into knowledge. By managing the information better and eliminating the irrelevant we can reduce the time it takes for human to find as they need to read. Text mining is a discovery through which we automatically extract the information from different written resources. Text mining also known as intelligent text analysis, text mining or knowledge discovery in text refers generally to the process of extracting interesting and non-trivial information and knowledge from unstructured text.

The Information Retrieval gives the subset of the overall information based on query. The problem of information overload is not solved here. Document Retrieval DR retrieves number of documents still beyond the capacity of human analysis, e.g. at the time of writing the query for information

retrieval in Google1 returned more than 30,100,000 results. Thus DR is not sufficient and we need a second level of abstraction to reduce this huge amount of data; the ability of summarization. This work tries to address this issue and proposes an automatic text summarization (TS) technique. Summarization is the process of reducing a large volume of information to a summary or abstract preserving only the most essential things. It produces a compressed version of overall document preserving the essential context. A TS system has to deal with natural language text and the complexities associated with natural language are inherited in the TS systems. Natural language text is unstructured and could be semantically ambiguous. Text Summarization is a very hard task as the computer must somehow understand what is important and what is not to be able to summarize. A TS system must interpret the contents of a text and preserve only most essential context. This involves extraction of syntactic and semantic information from the text and using this information to decide essentialness of the context. The following sub-section describes the need of TS systems with an example. According to Pooya Khosraviyan[14] human understand the contents, identifying the most important piece of information in the text to produce summary. In this work we present a text summarization technique based strategic approach which apply on the some feature contained in the sentences of the document. We ranked each sentences based on their feature and use manually summarized data for calculation of weight of each feature. We also use graph theoretic link reduction technique called threshold Scaling techniques. The text is represented as a graph with individual sentences as the nodes and lexical similarity between the sentences as the weights on the links. To calculate lexical similarity between the sentences it is necessary to represent the sentences as vectors of terms. Two sentences are more similar if they contain more common terms. In this work the features are the content words and the process of transformation from text to vectors is described in detail further on. The sentences (nodes in the graph) are then selected for inclusion in the final summary on the basis of their relative importance in the graph and feature score in the text.

II. TEXT SUMMARIZATION

Text summarization corresponds to the process in which a computer creates a compressed version of the original text (or a

collection of texts) still preserving most of the information present in the original text. This process can be seen as compression and it necessarily suffers from information loss. Simpler approaches were then explored that consist of extracting representative text-spans, using statistical techniques or the techniques based on surface domain-independent linguistic analyses. This is typically done by ranking document sentences and selecting those with higher score and minimum overlap. Thus a TS system must identify important parts and preserve them. What is important can depend upon the user needs or the purpose of the summary.

A. Classification

TS systems can be classified according to characteristics of many dimensions [18, 19]. Input: Characteristics of source text.

i) Source size: Single vs. Multi Document:

Single document, in such systems the summary is compressed version of only one text. A multi-document summary is one text that covers the content of more than one input text, and is usually used only when the input texts are thematically related.

ii) Specificity: Domain Specific vs. General:

When the input texts all related to a single domain, it may be appropriate to apply domain-specific summarization techniques, focus on specific content, and output specific formats, compared to the general case. A domain-specific summary derives from input text whose themes related to a single restricted domain. As such, it can assume less term ambiguity, idiosyncratic word and grammar usage, special formatting, etc., and can reflect them in the summary. A general-domain summary derives from input text in any domain, and can make no such assumptions.

iii) Genre and scale:

Typical input genres include newspaper articles, newspaper editorials or opinion pieces, novels, short stories, non-fiction books, progress reports, business reports, and so on. The scale may vary from book-length to paragraph-length. Different summarization techniques may apply to some genres and scales and not others.

B. Extractive Summarization

Sentence based extractive summarization techniques are commonly used in automatic summarization. The summary produced by the summarizer is a subset of the original text. Extractive summarizer picked out the most relevant sentences in the document with maintaining the low redundancy in the summary [2]. In this work the extraction unit is defined as a sentence. Sentences are well defined linguistic entities and have self contained meaning. So the aim of an extractive summarization system becomes, to identify the most important sentences in a text. The assumption behind such a system is that there exists a subset of sentences that present all the key points of the text. In this case the general framework of an extractive summarizer is shown in figure 1.

As it can be seen from figure 1, extractive summarization works by ranking individual sentences [4, 8, 12]. Most of the extractive summarization systems differ in this stage. A

sentence can be ranked using a clue indicating its significance in the text. There are various matrices for sentence selection from the text to produce summary [4]. It is a task of classification of sentence [19].

1. Sentence boundary discrimination
2. Building Vocabulary of the contents
3. Calculation of sentence importance (ranking)
4. Selection of ranked sentences

Figure1: Framework of extractive text summarization system.

III. LITERATURE REVIEW

This section reviews the previous work in the area of extractive text summarization. Extractive summarization systems can be divided into supervised and unsupervised techniques. Supervised techniques specified in [6, 19] are generally based on binary classification task where the sentences are classified as either to be included in the summary or not. The supervised techniques have two drawbacks. First, they need annotated corpora which are expensive as the texts need to be annotated manually. Second problem is that they are not portable. Once a classifier has been trained for one genre of documents (e.g. news articles or scientific documents) it cannot be used on the other genre without retraining. On the other hand the unsupervised techniques do not need annotated corpora (although annotations can be used to improve the performance) and are portable across genre. The following subsections review some approaches to extracting task.

Luhn's work exploiting frequent words:

H.P Luhn is the father of information retrieval. In his pioneering work [11] used simple statistical technique to develop an extractive text summarization system. Luhn used frequency of word distributions to identify important concepts, i.e. frequent words, in the text. As there could be uninformative words which are highly frequent (commonly known as stop words), he used upper and lower frequency bounds to look for informative frequent words. Then sentences were ranked according to the number of frequent words they contained. The criterion for sentence ranking was very simple and would read something like this:

If the text contains some words that are unusually frequent then the sentences containing those words are important. This quite simple technique which uses only high frequent words to calculate sentence ranking worked reasonably well and was modified by others to improve performance. Luhn provide a framework which can be used to measure various feature score for each text in the document. I used this approach with the weight of each term in the text instead of only frequency.

Edmundson's work exploiting cue phrases:

Luhn's work was followed by H. P. Edmundson [2] who explored the use of cue phrases, title words and location heuristic. Edmundson tried all the combinations and evaluated the system generated summaries with human produced extracts. The methods used include;

Cue method: Those containing cue words/phrases like conclusion, according to the study, hardly are given a higher weight than those not containing them [16]. The cue method used a cue dictionary which contained bonus words (positive weight), stigma words (negative words) and null words (equivalent to stop words).

Key method: A Key Glossary of words whose frequency of occurrence is above certain percentage of the total words was used. Statistically significant words are given higher scores. Score of sentence is then computed as the sum of the scores of its constituent words. In [5, 16] reports that he considered the words present in the sentences containing cue words, as significant words. Later the score of words is modified to be count of that word in the document. This is later made into a relative measure, and is modified to be the frequency of this word in the document.

Title method: Sentences containing title words are considered to have scored higher. Title words are those that are present in the title of the document, and headings and sub-headings. The first sentence in the document is often treated as Title [13].

Position method: The positive method assigns positive weight to headings, leading and concluding sentences in the paragraphs, and the sentences in the first and last paragraph as well.

Edmundson's work showed that the combination Cue+Title+Location produced best extracts followed by Cue+Title+Location+Key. The result that use of frequency did not lead to any improvement suggested two things: 1. it suggested the need for a different representation from word frequencies and 2. System time and memory can be saved by excluding word frequencies.

Salton's graph-based method:

Gerard Salton and co-workers explored a different idea of extractive summarization. Their system [17] identifies a set of sentences (paragraphs) which represents the document subject based on a graph based representation of the text. They proposed a technique uses undirected graphs with paragraphs as nodes and links representing similarity between paragraphs. Intra document links between passages were generated. This linkage pattern was then used to compute importance of a paragraph. The decision concerning whether the paragraph should be kept is determined by calculating number of links to other paragraphs. In other words, an important paragraph is assumed to be linked to many other paragraphs.

The system was evaluated on a set of 50 summaries by comparing them with human constructed extracts. The system's performance was fairly well.

Other graph theoretic techniques have been successfully applied to the task of extractive text summarization. In [3] authors proposed a system called LexRank which uses threshold-based link reduction as a basis of Markov random walk to compute sentence importance. In all those methods the text is represented in the form of a weighted graph with sentences as nodes and intra-sentence dissimilarity as link weights, which is the same for this work. This graph is then

used to identify sentence importance using various graph-theoretic algorithms.

Techniques mentioned so far fall under the general category of unsupervised techniques. To make the discussion about extractive summarization more complete the following subsection reviews the first supervised extraction system in [6] and a subsequent work in [19].

IV. DOCUMENT REPRESENTATION

Humans understand text as a natural language, i.e. by the meaning of the individual textual units and their relationship with each other. Natural language has no limits on the vocabulary and no complete set of rules to define its syntax. Moreover, the interpretation of the text is complex a process and involves cognitive dimensions. For a computer to understand natural language is still a far goal. Computers mostly rely on an abstract representation of the text described by the occurrence of words in the text. This is done under the reasonable assumption that the presence of words represents meaning. This involves processing of the textual information and converting them into a form which can be used by computers, typically tables.

A. PREPROCESSING

Before extracting feature it is necessary to normalize the document in a suitable manner so that we can extract only the textual data from the document whether the source is HTML file or Pdf file. The computation of feature is based on word level. The preprocessing work involves sentence marker, punctuation marker, stemming etc.

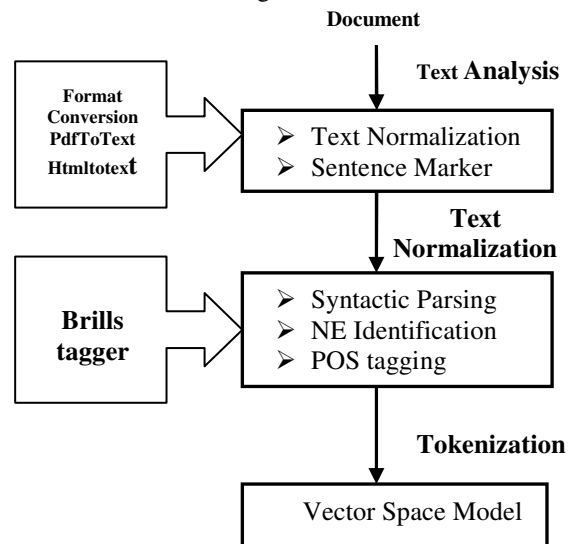


Figure 2: Preprocessing of text summarization

B. Text Analysis

As a part of summarization, we try to identify the important sentences which represent the document. This involves considerable amount of text analysis. We assume that the input document can be of any document format (ex. PDF, html ...), hence the system first applies document converters to extract the text from the input document. In our system we have used

document converters that could convert PDF, MS Word, post-script and HTML documents into text.

C. Text Normalization

The text normalization is a rule based component which removes the unimportant objects like figures, tables, identifies the headings and subheadings and handling of non-standard words like web URL's and emails and so on. The text is then divided into sentences for further processing.

D Sentence Boundary Marker

This module divides the document into sentences. In English two sentences are separated by using end-of-sentence punctuation marks, such as periods, question marks, and exclamation points (".", "?", "!"), is sufficient for marking the sentence boundaries. Exclamation point and question mark are somewhat less ambiguous. However, dot '.' in real text could be highly ambiguous and need not mean a sentence boundary always. The sentence marker considers the following ambiguities in marking the boundary of sentences.

Non standard word like web urls, emails, acronyms, and so on, will contain '.'

Every sentence starts with an uppercase letter

Document titles and subtitles can be written either in upper case or title case. For instance, the tiles like Mr., Ms., Prof. the symbol does not indicate sentence boundary.

E. Syntactic Parsing

This module analyzes the sentence structure with the help of available NLP tools such as Brills tagger, named entity extractor, etc. A named entity extractor can identify named entities (persons, locations and organizations), temporal expressions (dates and times) and certain types of numerical expressions from text. This named entity extractor uses both syntactic and contextual information. The context information is identified in the form of POS tags of the words and used in the named entity rules, some of these rules are general and while the rest are domain specific.

F. Tokenization or word parsing

The process by which the stream of characters is split into words (tokens) is called as tokenization. Tokens provide a basis for extracting higher level information from the unstructured text. Each token belongs to a type and thus could make repeated appearance in the text. As an example, text is a token that appeared twice in this paragraph. Tokenization is a non-trivial task for a computer due to lack of linguistic knowledge. So, certain word-boundary delimiters (e.g., space, tab) are used to separate the words. Certain characters are sometimes tokens and sometimes word boundary indicators. For instance, the characters - and: could be tokens or word-boundary delimiters depending on their context. of units: "Wb/m2" or "webers per square meter", not "webers/m2". Spell out units when they appear in text: "... a few henries", not "... a few H".

G. Vector-Space Model

After the work of preprocessing of the whole document, we get a dictionary consisting of unique set of tokens. This dictionary can be then used to describe the characteristic features of document.

In multi-document summarizer each document is converted into a numerical vector such that each cell of the vector is labeled with a word type in the dictionary and it contains its weight in the document. This weight is represented by binary value which denotes the presence or absence of the token in the document with the value 1 and 0 respectively. If the cell contains numerical value then it represents frequency (number of occurrences) of the term in the document. Thus the document is represented as an n-dimensional vector, one dimension for each possible term and hence the name [8]. We obtain a table in which the number of column is the total no of distinct word (term) and each rows correspond to the document.

It should be noted that the information about dependencies and relative position of the tokens in the document do not play any role in this representation, e.g. so "absence of light is darkness" is equivalent to "darkness is absence of light" in the vector-space model. Originally proposed by [17], vector space model is the frequently used numerical representation of text popularly used in information retrieval applications.

In single document summarization, the no of column is also representing the distinct word (term) and each rows representing the sentences. Each cell value represent whether the sentence containing that word (term) or not.

If each cell in a vector-space model is represented by term frequency (count of a type in the document) it is considered as local weighting of documents and is generally called as term frequency (tf) weighing. There are some words which occur very frequently than others. This is popularly known as the Zipf's law. This is because of the fact that there are not infinite numbers of words in a language. In 1949 in his landmark work Harvard linguist George K. Zipf argued that the word frequency follows power law distribution $f \propto r^{-a}$ with $a \approx 1$ [20], where f is the frequency of each word and r is its rank (higher frequency implies higher rank). This law, now known as Zipf's law, states that, frequency of a word is roughly inversely proportional to its rank.

To achieve this term frequency count can be weighed by the importance of a type in the whole collection. Such weighing is called as global weighing. One of such weighing schemes is called as inverse document frequency (idf). The motivation behind idf weighing is to reduce the importance of the words appearing in many documents and increasing importance of the words appearing in fewer documents. Then tf model when modified with idf results in the well-known tf-idf formulation [16]. The idf of a term t is calculated as following.

$$\text{idf}(t) = \log\left(\frac{N}{N_t}\right)$$

where N is the number of documents in the collection and N_t indicates number of documents containing the term t . The tf-idf measure combines the weight of each term in the sentence of the document. The term frequency, number of documents

and the number of documents in which the term is present and is calculated as;

$$W(t) = \text{tf-idf}(t) = \text{tf} * \text{idf}(t)$$

This vector space model provides a workspace through which we can compute various feature of each sentences.

Similarity Measures

Number of common words could be used as a measure of similarity between two texts. More sophisticated measures have been proposed which consider the number of words in common and number of words not in common and also lengths of the texts [10, 15]. Let us consider that, we want to measure similarity between two texts T1 and T2. The vocabulary consists of n terms, t1...tn. We use the notations tT1i and tT2i to represent the term occurrence in the text T1 and T2 respectively and can take either binary or real values.

Cosine coefficient

This is perhaps the most popular similarity measure. This measure calculates the cosine angle between two vectors in the high dimensional vector-space [1].

$$\text{Cosine}(T1, T2) = \frac{\sum_{t=1}^{t=n} W_t(T1) \cdot W_t(T2)}{\sqrt{\sum_{t=1}^{t=n} W_t^2(T1)} \sqrt{\sum_{t=1}^{t=n} W_t^2(T2)}}$$

This is an explicit measure of similarity. It considers each document as a vector starting at the origin and the similarity between the documents is measured as the cosine of the angle between the corresponding vectors.

The process of text summarization can be decomposed into three phases: analysis, transformation, and synthesis. The analysis phase analyzes the input text and selects a few salient features. The transformation phase transforms the results of analysis into a summary representation. Finally, the synthesis phase takes the summary representation, and produces an appropriate summary corresponding to users' needs. In the overall process, compression rate, which is defined as the ratio between the length of the summary and that of the original, is an important factor that influences the quality of the summary. As the compression rate decreases, the summary will be more concise; however, more information is lost. While the compression rate increases, the summary will be larger; relatively, more insignificant information is contained. In fact, when the compression rate is 5–30%, the quality of the summary is acceptable [5, 6].

In our proposed method of summarization each sentence is represented as a vector of feature score, and the document is represented as matrix. This matrix is multiplied with the weight matrix computed through manually summarized text corpus to get the score of each sentences. Then according to summary factor we select the sentences in descending order of their score in their order. In statistical method [6] was described by using a

Bayesian classifier to compute the probability that the sentence in the source document should be included in the summary. In [7, 8] there are various feature corresponding to the sentences measure the important of sentence in the text.

FEATURE DEFINITION

In this section we present various feature both for sentence level and word level which are used in calculating the importance or relevance of the sentences.

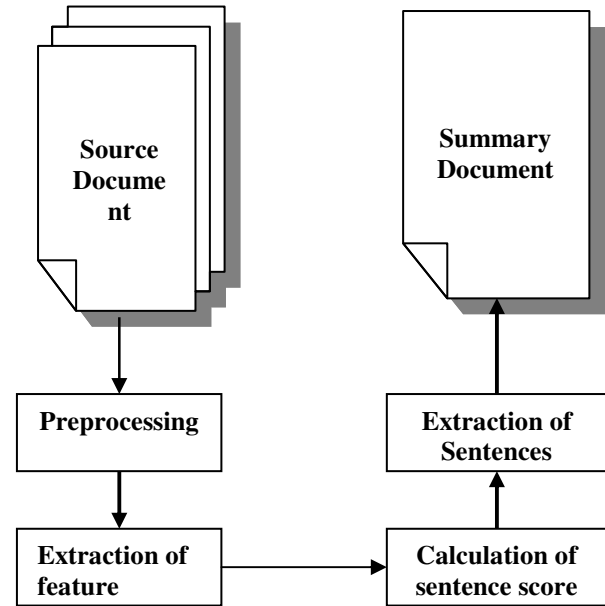


Figure 3: Proposed model of Automatic Text Summarization

F1: Sentence Position

We assume the first sentence of a paragraph is the most important. Therefore we rank a sentence in the paragraph according to their position. e.g. if there are 5 sentences in the paragraph then the 1st sentence have a score of 5/5, Then 2nd have score 4/5, 3rd have 3/5 and so on.

F2 = Positive keyword in the sentence

Positive keyword is the keyword frequently included in the summary. It can be calculated as follows:

$$\text{Score}_n(S) = \frac{1}{\text{Length}(s)} \sum_{i=1}^n \text{tf}_i * P$$

$$\text{Where } P = \left| \frac{\text{No of Keywords in the sentence}}{\text{No of Keywords in the paragraph}} \right|$$

tf_i is the occurrence or frequency of ith term in the sentence, which probably is a keyword.

F3: Sentence Relative Length

This feature is useful to filter out short sentences such as datelines and author name commonly found in news articles. The short sentences are not expected to belong in the summary. We use length of the sentences, which is the ratio of the number of word occurring in the sentence over number of word in the longest sentence in the document.

$$\text{Score}_{f3}(S) = \left| \frac{\text{No of words occurring in Sentence S}}{\text{No of words occurring in longest sentence}} \right|$$

F4: Sentence resemblance to title

It is the measure of vocabulary overlap between this sentence and the document title, generally the first sentence in the document is probably the title of the document. It is calculated as

$$\text{Score}_{f4}(S) = \left| \frac{\text{Keyword in } S \cap \text{Key word in title}}{\text{Keyword in } S \cup \text{Key word in title}} \right|$$

F5: Sentence inclusion of name entity (Proper noun)

Usually the sentence that contains more proper nouns is an important one and it is most probably included in the summary. Proper noun gives the literature of contents.

$$\text{Score}_{f5}(S) = \left| \frac{\text{Number of proper noun in S}}{\text{Length of S}} \right|$$

F6: Sentence inclusion of numerical data

Sentences that contain numerical data are more important than rest of sentences and are probably included in the summary.

$$\text{Score}_{f6}(S) = \left| \frac{\text{Number of numerical data in S}}{\text{Length of S}} \right|$$

F7: Term Weight

The frequency of term occurrence within a document has often been used for calculating the importance of sentence. The score of sentence can be calculated as the sum of the score of word in the sentence.

The score or weight w_i of i th term or word can be calculated by traditional tf-idf method discuss in previous section [16].

$$\text{Score}_{f7}(S) = \frac{\sum_{i=1}^{i=k} W_i(S)}{\text{Max} \left(\sum_{i=1}^{i=k} W_i(S^N) \right)}$$

F8: Sentence similarity with other sentence

This feature measures the similarity between sentence S and each other sentences. It measures how much vocabulary overlap between this sentence and other sentences in the document. It is computed by cosine similarity measure with resulting between 0 and 1 [1]. The score of this feature for a sentence S is obtained by computing the ratio of similarity of sentence S with each other sentence over the maximum similarity between two sentences.

$$\text{Score}_{f8}(S) = \frac{\sum \text{Sim}(S, S_j)}{\text{MAX}(\sum \text{Sim}(S_i, S_j))}$$

Where $\text{Sim}(S_i, S_j)$ is cosine similarity between sentence S_i, S_j define previously

F9: Bushy path of the Sentence or node Sentence centrality

It has an overlapping vocal bury with several sentences it is defined as the number of links connected it to other sentences (node) on similarity graph. Highly busy node is linked to the number of other nodes. The busy path is calculated as follow:

$$\text{Score}_{f9}(S) = \frac{\# \text{ (branches connected to sentence (node) S)}}{\text{Max Degree in the scaled similarity graph}}$$

The Automatic method which is used to determine whether there is a link between two sentences in the similarity graph. The weight of link measure the strength of similarity which is measured previously, for computing the busy path we use scaling techniques which preserve only critical links.

A network in general represents concepts as nodes and links between concepts as relations with weights indicating strength of the relations. The hidden or latent structure underlying raw data, a fully connected network, can be uncovered by preserving only critical links. The aim of a scaling algorithm is to prune a dense network in order to reveal the latent structure underlying the data which is not visible in the raw data. Such scaling is obtained by generating an induced sub graph. There are two link-reduction approaches: threshold-based and topology-based. In threshold-based approach elimination of a link is solely decided depending upon whether its weight exceeds some threshold. On the other hand, a topology-based approach eliminates a link considering topological properties of the network. Therefore a topology-based approach preserves intrinsic network properties reliably. We have used a threshold based approach with a threshold of 0.04 to discard branches among nodes that similarity less than 0.04.

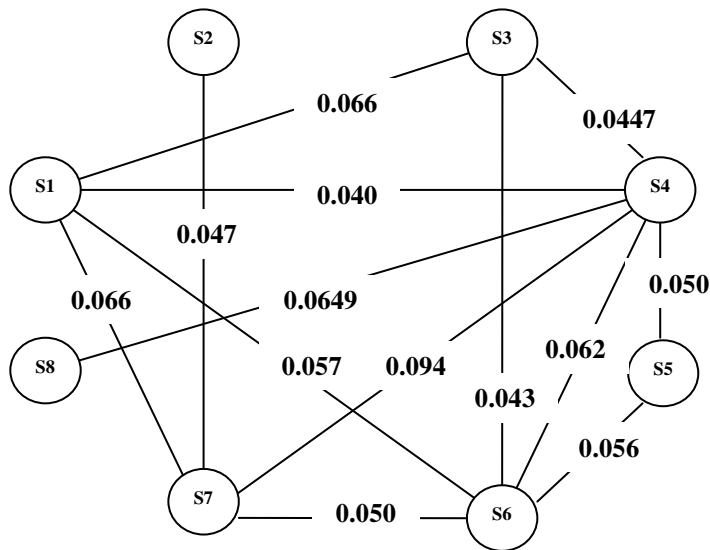


Figure 4: Scaled network graph with threshold of 0.04.

All the sentences are ranked by calculating various feature score for all sentences and according to the compression rate they selected for inclusion in summary in descending order of their rank in the order of their appearance.

Table 1: Feature Score and rank of the all sentences

Sentence No / Feature Score	1	2	3	4	5	6	7	8
Sentence Position	0.710	0.650	0.570	0.000	1.000	0.000	0.000	0.000
Positive Keyword	0.370	0.770	0.150	0.230	0.130	0.530	0.510	1.000
Sentence Relative Length	1.000	0.870	0.750	0.620	0.500	0.370	0.250	0.120
Proper Noun	0.780	0.850	0.970	0.950	0.560	0.680	1.000	0.430
Numerical Term	0.830	0.750	0.660	0.660	0.580	0.750	1.000	0.500
Term Weight	0.840	0.660	0.790	0.700	0.460	0.600	1.000	0.730
Similarity with others	0.714	0.889	1.000	0.898	0.618	0.647	0.994	0.476
Busy Path	0.667	0.167	0.500	1.000	0.333	0.833	0.667	0.167
Final Score	5.911	5.606	5.390	5.058	4.181	4.410	5.421	3.422
Rank	1	2	4	5	7	6	3	8

V. RESULT

Most of the summarization systems developed so far is for news articles. There are two major reasons for this: news articles are readily available in electronic format and also huge amount of news articles are produced every day. One interesting aspect of news articles is that they are written in such a way that usually most important parts are at the beginning of the text. So a very simple system that takes the required amount of leading text produces acceptable summaries. But this makes it very hard to develop methods that can produce better summaries.

Summary Evaluation:

The quality of summary is varies from human to human. The summary produced by human is to select the most relevant

sentence from a given document. This is different from different people. This makes the evaluation of task of automatic generated summaries is difficult and there is no standard available.

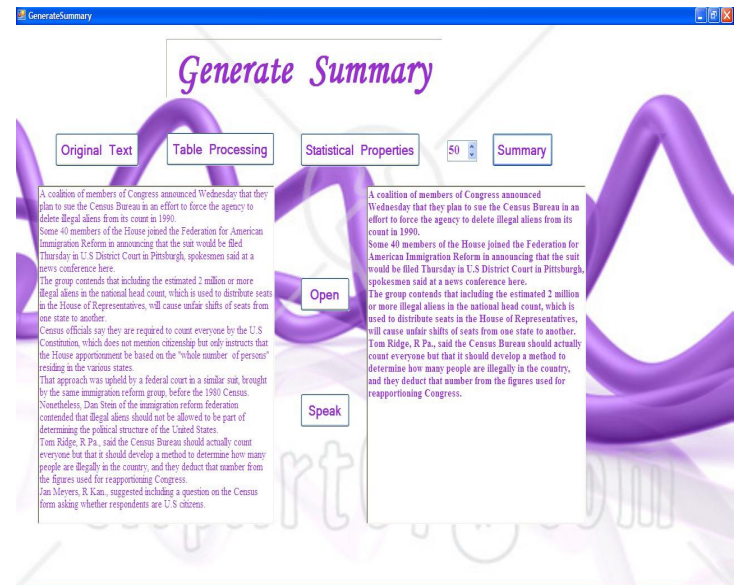


Figure 5: snapshot of generated summary.

There are some measures which quantify the quality of summaries produced. It is classified into two types.

- Intrinsic evaluation is a method which measures the quality of the summary as output.
- Extrinsic evaluation is a method which measures the quality of output summary in the form of its assistance in another task.

My work uses intrinsic evaluation. Most of the existing summary evaluation techniques are intrinsic in nature. Typically the system output is compared with ideal summary created by human evaluators. Since a summary is subjective often more than one ideal summary are used to get a better evaluation. Many researchers have used this kind of evaluation [2, 6, 16]. Edmundson proposed a method for measuring quality of extracts. In his method extracts sentences are compared with the sentences hand-picked by human judges. The quality of an automatically summary is measured by computing number of sentences common between the automatically generated summary and the human summary. Although this method is widely used it involves a lot of manual work and for the same reason it is inapplicable to large scale evaluations. Recently Lin and Hovy proposed automatic measures of summary evaluation called ROUGE [9].

We use an intrinsic evaluation to judge the quality of a summary based on the coverage between it and the manual summary. We measure the system performance in terms of precision and Recall from the following formula:

$$\text{Recall} = \frac{|\text{S} \cap \text{T}|}{|\text{T}|}$$

$$\text{Precision} = \frac{|\text{S} \cap \text{T}|}{|\text{S}|}$$

Where T is the manual summary and S is the machine-generated summary.

Generally in information retrieval tasks increase in precision causes decrease in recall and vice versa. That means they are inversely related. F measure is used to combine precision and recall. An ideal system should have both high precision and high recall. But as maximum of both cannot be achieved they are combined into F measure to get an idea about general behavior of the system. F measure is defined as:

$$F1 = \frac{(\alpha+1) \text{ Precision} * \text{ Recall}}{\alpha * (\text{ Precision} + \text{ Recall})}$$

$$F2 = \frac{2 * \text{ Precision} * \text{ Recall}}{\text{ Precision} + \text{ Recall}}$$

In F1 measure recall and precision are given equal importance. Other measures giving different importance to precision and recall are also possible, for example, F2 measure gives twice as much weight to recall than to precision.

Table2 shows the evaluation of the summary produced by our tool ATS, which is compared with the summary produced by Microsoft Word. The precision, Recall and harmonic mean of Precision & Recall is computed for ten News Articles from www.paperarticles.com.

Table 2: Performance evaluation of 10 news article by Precision (P), Recall(R) and F2 measure for different compression rate.

Compression Rate		corpus1	corpus2	corpus3	corpus4	corpus5	corpus6	corpus7	corpus8	corpus9	corpus10
10%	P	1.00	0.14	0.14	0.25	0.17	0.00	0.17	0.33	0.17	0.17
	R	1.00	0.14	0.11	0.20	0.20	0.00	0.13	0.33	0.17	0.20
	F2	1.00	0.14	0.13	0.22	0.18	0.00	0.14	0.33	0.17	0.18
20%	P	0.50	0.29	0.33	0.33	0.36	0.20	0.25	0.33	0.31	0.25
	R	0.50	0.29	0.31	0.38	0.33	0.21	0.20	0.40	0.40	0.30
	F2	0.50	0.29	0.32	0.35	0.35	0.21	0.22	0.36	0.35	0.27
30%	P	0.50	0.32	0.36	0.43	0.47	0.27	0.33	0.33	0.26	0.42
	R	0.33	0.35	0.36	0.46	0.53	0.32	0.29	0.38	0.31	0.47
	F2	0.40	0.33	0.36	0.44	0.50	0.29	0.31	0.35	0.29	0.44
40%	P	0.33	0.47	0.33	0.44	0.61	0.50	0.33	0.42	0.42	0.36
	R	0.25	0.54	0.33	0.44	0.70	0.60	0.30	0.45	0.52	0.43
	F2	0.29	0.50	0.33	0.44	0.65	0.55	0.31	0.43	0.47	0.39
50%	P	0.50	0.53	0.51	0.50	0.64	0.62	0.53	0.57	0.50	0.55
	R	0.50	0.56	0.54	0.50	0.78	0.74	0.52	0.57	0.59	0.65
	F2	0.50	0.55	0.53	0.50	0.71	0.68	0.52	0.57	0.54	0.60

VI CONCLUSION & FUTURE WORK

This work presents a new extractive text summarization technique, for single documents based on Feature Extraction. Extractive text summarization works by selecting a subset of important sentences from the original document. We used text processing approaches as opposed to semantic approaches related to natural language. To calculate the similarity we use the well known tf*idf model of document representation. Such graphical representation gives us a way to calculate sentence importance. The centrality reveals the relative importance of a sentence in the text. Our work does not need natural language processing resources apart from a word and sentence boundary parsers and a stemmer (optional). Thus the method can be extended to other languages with little modifications.

In our system we have come up with arbitrary weights by trial and error method. We plan to implement machine learning techniques to learn these weights automatically from training data. We would like to use NLP tools such as word sense disambiguation and co-reference resolution module to obtain precise weights for the sentences in the document we also plan to extend this system to perform deeper semantic analysis of the text and add more feature to our ranking function. We would like to extend this system for multi document summarization. Semantic information such as word sense can be utilized. Same word can mean different things in different contexts. Use of word sense information can lead to better similarity calculations. Same word can be used in different senses in different context. So using the correct word sense can lead to better similarity measurements. A more sophisticated representation that single words can be explored. A first step towards this aim could be use of multi-word units. Multi-word units can be recognized using statistical techniques. Also syntactic information such as Part-of- Speech (POS) tags might help to improve performance of the extraction algorithm.

REFERENCES

- [1] Amy J.C Trappey, Charles V . Trappey, "An R&D Knowledge Management method of patent document summarization", Industrial Management & Data System, vol 108 pp -245-257,2008.
- [2] Edmundson, H. P(1969) "New method in automatic Extracting" journal of ACM 1969, 16(2): 264-285.
- [3] Erkan G., and Radev, D. R., "LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization", J. Artif. Intell. Res. (JAIR), 22, pp. 457-479, 2004.
- [4] Goldstein, J., Kantrowitz, M., Mittal, V., Carbonell, J., 1999. Summarizing text documents: sentence selection and evaluation metrics. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99), Berkeley, CA, USA, pp. 121-128.
- [5] Hahn, U., Mani, I., 2000. The challenges of automatic summarization. IEEE-Computer 33 (11), 29-36.
- [6] Kupiec J, Pedersen J.O, Chen F.A(1995) trainable document summarizer, In Proceedings of the 18th ACM-SIGIR Conference, Association of Computing Machinery
- [7] Ladda Suanmali, Naomie Salim, Fuzzy Logic Based method for improving text summarization, International journal of computer science and information security, vol. 2, No. 1, 2009.
- [8] Lee, D.L Chuang,H. &seamons K, "Document ranking and vector space model, software, IEEE, 14(2),67-75, 1997.

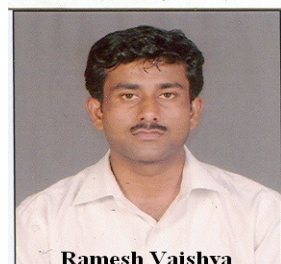
- [9] Lin, C. -Y. (2004). ROUGE: A package for automatic evaluation summaries. In Proceedings of the workshop on text summarization branches out, 25–26 July (pp.74–81) Barcelona, Spain.
- [10] Liu, X., Zhou, & Y., Zheng, R. (2007). Sentence similarity based on dynamic time warping. In Proceedings of the first international conference on semantic computing (ICSC 2007), 17–19 September (pp. 250–256) Irvine, USA.
- [11] Luhn, H. P. 1958. The automatic creation of literature abstracts. IBM Journal of Research Development, 2(2):159-165
- [12] M. A Fattah , Fuji Ren,(2008)“Automatic Summarization ”, in proceeding of word Academy of Science, Engineering and Technology volume 27 pp 192-195.
- [13] M. A Fattah , Fuji Ren,“GA, MR, FFNN, PNN and GMM based models for automatic text summarization”, Computer Speech and Language 23 (2009) 126–144, www.elsevier.com/locate/csl.
- [14] Pooya khosraviyan dehkordi, dr. Farshad& dr. Hamid khosravi,” Text Summarization Based on Genetic Programming”, International Journal of Computing and ICT Research, Vol. 3, No. 1, June 2009.
- [15] Ramiz M. Aliguliyev, A new sentence similarity measure and sentence based extractive technique for automatic text summarization, Expert Systems with Applications 36 (2009) 7764–7772, journal homepage: www.elsevier.com/locate/eswa.
- [16] Salton, G. and Buckley, C. Term-weighting approaches in automatic text retrieval. Information processing and management, 1998, 24(5):513-523.
- [17] Salton, G., Wong, A., and Yang, C., A vector space model for automatic indexing, Communications of the ACM, 18, pp. 613-620, 1975.
- [18] Spark Jones, Karen. 1999. Automatic summarizing: Factors and directions. In I. Mani and M.T. Maybury, editors, Advances in Automatic Text Summarization. MIT Press, Cambridge, pages 1-13.
- [19] Teufel, S.H., Moens, M., 1997. Sentence extraction as a classification task. In: Proceedings of the ACL’97/EACL’97 Workshop on Intelligent Scalable Text Summarization, Madrid, Spain, pp. 58–65.
- [20] Zipf , G.K., Human Behaviour and the Principle of Least-Effort, Addison-Wesley, Cambridge MA, 1949.

AUTHORS PROFILE



¹**Dr. Surya Prakash Tripathi** is currently working as a Associate Professor in Department of Computer Science & Engineering at I.E.T Lucknow. He has twenty three year teaching experience in computer science & engineering field. He has published number of papers in referred National journals

His teaching areas are: Software Engineering, Database Management, Operating System, Data mining and Computer Network



Ramesh Vaishya

²**Ramesh Vaishya** is currently working as a Senior Lecturer in Department of Computer Science & Engineering at B.B.D.N.I.T.M (Babu Banarsi Das National Institute of Technology & Management), Lucknow. He has total Eight year teaching experience in computer science & engineering field.

His teaching areas are: Database Management, Data Structure, Design & Analysis of Algorithm, Compiler Design and Object Oriented System.

* **Corresponding Author**

Golomb Ruler Sequences Optimization: A BBO Approach

Shonak Bansal

Department of Electronics
& Communications,
Maharishi Markandeshwar
University, Mullana,
Haryana, INDIA
shonakk@gmail.com

Shakti Kumar

Computational Intelligence
Laboratory
Institute of Science and
Technology, Klawad,
Haryana, INDIA
shaktik@gmail.com

Himanshu Sharma

Department of Electronics
& Communications,
Maharishi Markandeshwar
University, Mullana,
Haryana, INDIA
himanshu.zte@gmail.com

Parvinder Bhalla

Computational Intelligence
Laboratory
Institute of Science and
Technology, Klawad,
Haryana, INDIA
parvinderbhalla@gmail.com

Abstract— The Four Wave Mixing (FWM) crosstalk with equally spaced channels from each other is the dominant nonlinear effect in long haul, repeaterless, wavelength division multiplexing (WDM) lightwave fiber optical communication systems. To reduce FWM crosstalk in optical communication systems, unequally spaced channel allocation is used. One of the unequal bandwidth channel allocation technique is designed by using the concept of Golomb Ruler. It allows the gradual computation of a channel allocation set to result in an optimal point where degradation caused by inter-channel interference (ICI) and FWM is minimal. In this paper a new Soft Computing approach called Biogeography Based Optimization (BBO) for the generation and optimization of Golomb Ruler sequences is applied. It has been observed that BBO approach perform better than the two other existing classical methods i.e. Extended Quadratic Congruence (EQC) and Search Algorithm (SA).

Keywords— Four wave mixing, Optimal Golomb Ruler, Soft Computing, Biogeography Based Optimization.

I. INTRODUCTION

In conventional wavelength division multiplexing systems, channels are usually assigned with center frequencies (or wavelength) equally spaced from each other. Due to equal spacing among the channels there is very high probability that noise signals (such as FWM signals) may fall into the WDM channels, resulting in severe crosstalk [1].

FWM crosstalk is the main source of performance degradation in all WDM systems. Performance can be substantially improved if FWM generation at the channel frequencies is avoided. It is therefore important to develop algorithms to allocate the channel frequencies in order to minimize the FWM effect. The efficiency of FWM depends on the channel spacing and fiber dispersion [2], [3]. If the frequency separation of any two channels of a WDM system is different from that of any other pair of channels, no FWM signals will be generated at any of the channel frequencies. This suppresses FWM crosstalk [4] – [7]. Thus, the use of proper unequal channel spacing keeps FWM signals from coherently interfering with the desired signals.

In order to reduce the FWM crosstalk effects in WDM systems, several unequally spaced channel allocation (USCA) techniques have been studied in literature [1], [8] – [14]. An optimum USCA (O-USCA) technique ensures that no FWM signals will ever be generated at any of the channel frequencies if the frequency separation of any two channels is

different from any other pair of channels in a minimum operating bandwidth [11].

Forghieri et al. [6] treated the “channel-allocation” design as an integer linear programming (ILP) problem by dividing the total available bandwidth into equal frequency slots. But the ILP problem was NP-complete and no general or efficient method was known to solve the problem. So optimum solutions (i.e., channel locations) were obtained only with an exhaustive computer search [1].

However, the techniques [8] – [14] have the drawback of increased bandwidth requirement as compared to equally spaced channel allocation. This is due to the constraint of the minimum channel spacing between each channel and that the difference in the channel spacing between any two channels is assigned to be distinct. As the number of channel increases, the bandwidth for the unequally spaced channel allocation methods increases in proportion [4].

This paper proposes a method for finding the solutions to channel allocation problem by using the concept of Optimal Golomb Rulers (OGR) [7], [15] – [17]. This method for channel allocation achieves reduction in FWM effect with the WDM systems without inducing additional cost in terms of bandwidth. This technique allows the gradual computation of a channel allocation set to result in an optimal point where degradation caused by inter-channel interference (ICI) and FWM is minimal [4], [16].

Much effort has been made to compute short or dense RUs and to prove them optimal. Golomb Rulers represent a class of problems known as NP – complete [18]. Unlike the traveling salesman problem (TSP), which may be classified as a *complete ordered set*, the Golomb Ruler may be classified as an *incomplete ordered set*. The exhaustive search [19], [20] of such problems is impossible for higher order models. As another mark is added to the ruler, the time required to search the permutations and to test the ruler becomes exponentially greater. The success of Soft Computing approaches such as Genetic Algorithms (GAs) [21] – [23] in finding relatively good solutions to NP – complete problems provides a good starting point for methods of finding Optimal Golomb Ruler sequences. Hence, soft computing approaches seem to be very effective solutions for the NP – complete problems. No doubt, these approaches do not give the exact or best solutions but reasonably good solutions are available at given cost. In this paper, a novel optimization algorithm based on the theory of biogeography of species called Biogeography Based

Optimization (BBO) is being applied to generate the optimal Golomb Ruler sequences for various marks.

The remainder of this paper is organized as follows: Section II introduces the concept of Golomb Rulers. Section III presents the problem formulation. Section IV describes a brief introduction about BBO and steps to generate the Golomb Ruler sequences by using this soft computing approach. Section V provides simulation results comparing with conventional classical approaches of generating unequal channel spacing i.e. Extended Quadratic Congruence (EQC) and Search Algorithm (SA). Section VI presents some concluding remarks.

II. GOLOMB RULERS

The idea of 'Golomb Rulers' was first introduced by W.C. Babcock [7] in 1952, and further derived in 1977 from the relevant work by Professor Solomon W. Golomb [15], a professor of Mathematics and Electrical Engineering at the University of Southern California. According to Colannino [24] and Dimitromanolakis [25], W. C. Babcock [7] first discovered Golomb Rulers up to 10- marks, while analyzing positioning of radio channels in the frequency spectrum. He investigated inter-modulation distortion appearing in consecutive radio bands and observed that when positioning each pair of channels at a distinct distance, then third order distortion was eliminated and fifth order distortion was lessened greatly. According to William T. Rankin [26], all of rulers' upto eight are optimum, the nine and ten mark rulers that W. C. Babcock presents are near optimum.

The term 'Golomb Ruler' refers to a set of non-negative integers such that no distinct pairs of numbers from the set have the same difference [27]. These numbers are referred to as marks [15], [21], [28] and correspond to positions on a linear scale. The difference between the values of any two marks is called the distance between those marks. The difference between the largest and smallest number is referred to as the length of the ruler. The number of marks on a ruler is sometimes referred to as the size of the ruler. Unlike usual rulers, Golomb Rulers measure more discrete lengths than the number of marks they carry. Normally the first mark of the ruler [15], [16], [29] is set on position 0. Since the difference between any two numbers is distinct, the new FWM frequencies generated would not fall into the one already assigned for the carrier channels. Golomb Rulers are not redundant as they do not measure the same distance twice [29].

Figure 1 shows an example of Golomb Ruler. The distance between each pair of marks is also shown in the figure [21].

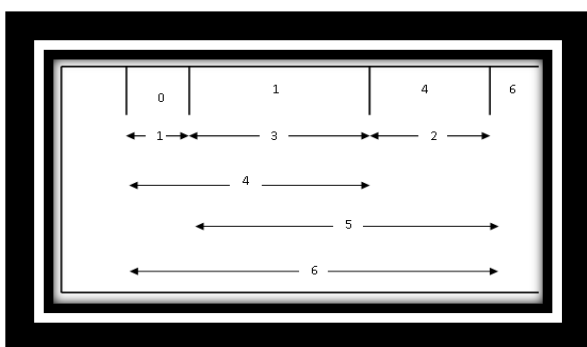


Figure 1. A Golomb Ruler with 4 Marks and Length 6

The particularity of Golomb Rulers is that all differences between pairs of marks are unique [29], [30]. Although the

definition of a Golomb Ruler does not place any restriction on the length of the ruler, researchers are usually interested in rulers with minimum length.

A perfect Golomb Ruler measures all the integer distances from 0 to L, where L is the length of the ruler [18], [21], [22]. In other words, the difference triangle of a perfect Golomb Ruler contains all numbers between one and the length of the ruler. The length [31] of an n - mark perfect Golomb Ruler is $\frac{1}{2}n(n-1)$.

For example, as shown in Figure 2 the set (0, 1, 3, 7) is a non optimal 4-mark Golomb Ruler since its differences are (1 = 1 - 0, 2 = 3 - 1, 3 = 3 - 0, 4 = 7 - 3, 6 = 7 - 1, 7 = 7 - 0), all of which are distinct. As from the differences it is clear that the number 5 is missing so it is not a perfect Golomb Ruler sequence.

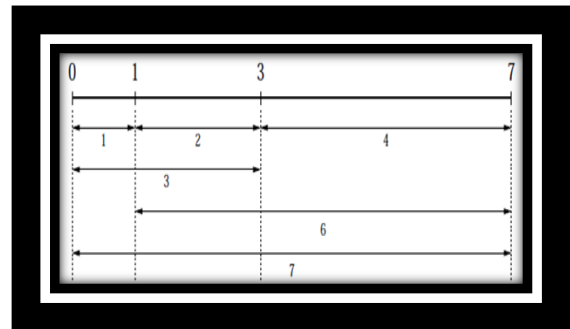


Figure 2. A Non Optimal Golomb Ruler of 4-Marks and Length 7

However, the unique optimal Golomb 4-mark ruler is (0, 1, 4, 6), which measures the distances (1, 2, 3, 4, 5, 6) (and is therefore also a perfect ruler) as shown in Figure 1.

An Optimal Golomb Ruler is defined as the shortest length ruler for a given number of marks [21], [32]. There can be multiple different OGRs for a specific number of marks.

The OGRs are used in a variety of real - world applications including Communications and Radio Astronomy, X-Ray Crystallography, Coding Theory, Linear Arrays, Computer Communication Network, PPM Communications, circuit layout, geographical mapping and Self-Orthogonal Codes [7], [15], [21], [22], [26].

An n - mark Golomb Ruler is a set of n distinct nonnegative integers (a_1, a_2, \dots, a_n) , called "marks," such that the positive differences $|a_i - a_j|$, computed over all possible pairs of different integers $i, j = 1, 2, \dots, n$ with $i \neq j$ are distinct [20]. Let a_n be the largest integer in an n - mark Golomb Ruler [33]. Then an n - mark Golomb Ruler $(0, \dots, a_n)$ is said to be optimal if and only if

1. There exists no other n -mark Golomb Rulers having smaller largest mark a_n , and
2. The ruler is written in canonical form as the "smaller" of the equivalent rulers $(0, a_2, \dots, a_n)$ and $(0, \dots, a_n - a_2, a_n)$, where "smaller" means the first differing entry is less than the corresponding entry in the other ruler.

In such a case, a_n is called the *length* of the optimal n - mark ruler.

Various classical methods are proposed in [1], [8] - [14] to generate the OGRs. The soft computing methods that employ genetic algorithm (GA) based methods [21] - [23] could be found in literature. This paper proposes a new soft computing technique based on the mathematics of biogeography to

generate Golomb Ruler sequences, i.e., biogeography based optimization algorithm and its performance comparison with existing classical methods that employ EQC and SA [1], [13], [21].

III. PROBLEM FORMULATION

If the spacing between any pair of channels is denoted as CS and the total number of channels is N , then the objective is to minimize the length of the ruler denoted as RL , which is given by the equation (1):

$$RL = \sum_{i=1}^N (CS)_i \quad (1)$$

subject to $(CS)_i \neq (CS)_j$

where $i, j = 1, 2, \dots, n$ with $i \neq j$ are distinct.

If each individual element is a Golomb Ruler, the sum of all elements of an individual forms the bandwidth of the channels. Thus, if an individual element is denoted as IE and the total number of elements is M , then the second objective is to minimize the bandwidth (BW), which is given by the equation (2):

$$BW = \sum_{i=1}^M (IE)_i \quad (2)$$

subject to $(IE)_i \neq (IE)_j$

where $i, j = 1, 2, \dots, n$ with $i \neq j$ are distinct.

IV. SOFT COMPUTING APPROACH

In this section, the capabilities of a new technique based on the mathematics of biogeography called BBO for the generation of optimal Golomb Ruler sequences will be discussed.

A. Biogeography Based Optimization

Biogeography Based Optimization is a population-based evolutionary algorithm (EA) developed for global optimization. It is based on the mathematics of biogeography. It is a new kind of optimization algorithm which is inspired by the science of Biogeography. It mimics the migration strategy of animals to solve the problem of optimization [34] – [39]. Biogeography is the study of the geographical distribution of biological organisms. Biogeography theory proposes that the number of species found on habitat is mainly determined by immigration and emigration. Immigration is the arrival of new species into a habitat, while emigration is the act of leaving one's native region. The science of biogeography can be traced to the work of nineteenth century naturalists such as Alfred Wallace [40] and Charles Darwin [41].

In BBO, problem solutions are represented as islands and the sharing of features between solutions is represented as emigration and immigration. An island is any habitat that is geographically isolated from other habitats [42].

The idea of BBO was first presented by Dan Simon in December 2008 and is an example of how a natural process can be modeled to solve general optimization problems [43]. This is similar to what has occurred in the past few decades with Genetic Algorithms (GAs), Artificial Neural Networks (ANNs), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and other areas of computer intelligence. Biogeography is nature's way of distributing species, and is analogous to general problem solving. Suppose that there are some problems and that a certain number of candidate solutions are there. A good solution is

analogous to an island with a high HSI (Habitat suitability index), and a poor solution is like an island with a low HSI.

Features that correlate with HSI include factors such as distance to the nearest neighboring habitat, climate, rainfall, plant and animal diversity, diversity of topographic features, land area, human activity, and temperature [39]. The variables that characterize habitability are called suitability index variables (SIVs). High HSI solutions are more likely to share their features with other solutions, and HSI solutions are more likely to accept shared features from other solutions [43] – [45]. As with every other evolutionary algorithm, each solution might also have some probability of mutation, although mutation is not an essential feature of BBO the improvement of solutions is obtained by perturbing the solution after the migration operation [46].

1) BBO Algorithm to Generate Optimal Golomb Ruler Sequences

The basic structure of BBO algorithm to generate OGR sequences is as follows:

1. Initialize the BBO parameters: maximum species count i.e. population size S_{max} , the maximum migration rates E and I , the maximum mutation rate m_{max} , an elitism parameter and the number of iterations.
2. Initialize the number of channels (or marks) 'N' and the upper bound on the length of the ruler.
3. Initialize a random set of habitats (integer population), each habitat corresponding to a potential solution to the given problem. The number of integers in each habitat being equal to the number of channels or mark input by the user.
4. Check the golombness of each habitat. If it satisfies the conditions for Golomb Ruler sequence, retain that habitat; if it does not, delete that particular habitat from the population generated from the step 3.
5. For each habitat, map the HSI (Total Bandwidth) to the number of species S , the immigration rate λ , and the emigration rate μ .
6. Probabilistically use immigration and emigration to modify each non-elite habitat, then recompute each HSI.
7. For each habitat, update the probability of its species count given by equation (3). Then, mutate each non-elite habitat based on its probability, check golombness of each habitat again and then recompute each HSI.

$$\dot{P}_s = \begin{cases} -(\lambda_s + \mu_s) P_s + \mu_{s+1} P_{s+1}, & S = 0 \\ -(\lambda_s + \mu_s) P_s + \lambda_{s-1} P_{s-1} + \mu_{s+1} P_{s+1}, & 1 \leq S \leq S_{max} - 1 \\ -(\lambda_s + \mu_s) P_s + \lambda_{s-1} P_{s-1}, & S = S_{max} \end{cases} \quad (3)$$

where λ_s and μ_s are the immigration and emigration rates, when there are S species in the habitat.

8. Is acceptable solution found? If yes then go to Step 10.
9. Number of iterations over? If no then go to Step 3 for the next iteration.
10. Stop

V. SIMULATION RESULTS AND DISCUSSION

In this section, the performance of BBO approach to generate unequal channel spacing sequences called Golomb Rulers and its comparison with known OGR [24], [33], [47], [48] and conventional classical methods of generating unequal channel spacing i.e. Extended Quadratic Congruence and Search Algorithm [1], [13], [21] is discussed. The algorithm to generate optimal Golomb Ruler sequences has been written and tested in Matlab – 7 [49] language under Windows 7 operating system. This algorithm has been executed on Laptop with Intel core 2 Duo processor with a RAM of 3 Gb.

A. Simulation Parameters for BBO Algorithm

To get optimal solution after a number of careful experimentation, following optimum values of BBO parameters have finally been settled as shown in Table I.

TABLE I. SIMULATION PARAMETERS FOR BBO ALGORITHM

Parameter	Value
Habitat modification probability (P_{modify})	1
Lower bounds of immigration probability per gene (λ_{Lower})	0
Upper bounds of immigration probability per gene (λ_{Upper})	1
Step size (dt) for numerical integration of probabilities	1
Maximum immigration (I) rates for each island	1
Maximum emigration (E) rates for each island	1
Mutation probability (P_{mutate})	0.05
Elitism (keep) per generation	2

B. Sequences

The optimum Golomb Ruler sequences generated by Biogeography Based Optimization algorithm are shown in Appendix – A for different values of marks. It has been verified that all the generated sequences are Golomb Rulers.

C. Influence of Increasing Iterations on Total Bandwidth

As the number of iterations increases, the total bandwidth of the sequence tends to decrease; it means that the rulers reach their optimum values after a certain number of iterations. This is the point where the results are optimum and no further improvement is seen, that is, we are approaching towards the optimal solution. This can be seen in tabular form for BBO in Table IV for various marks and graphically in Figure 3.

In Table II, 'N' is the number of marks (called channels) in Golomb Ruler sequences. It is noted that the iterations has little effect for low value marks say for N = 3, 4 and 5 so they are not shown in Figure 3. But for higher order marks, the generations has a great effect on the total bandwidth i.e. bandwidth gets optimized after a certain numbers of

iterations. By carefully observation, the paper fixed the iterations of 5000 for BBO algorithm.

D. Influence of Population Size on the Performance of BBO Approach

In this subsection, the influence of population size (Popsiz) on the performance of soft computing approach (BBO) for various values of marks is investigated. Increasing the population size will increase the diversity of possible solutions, and promote the exploration of the search space. But the choice of the best population size of BBO is problem-specific [39]. In this experiment, all the parameter settings for BBO are same as mentioned in above subsection V–A except for population size. Table III shows the influence of population size on total bandwidth and ruler length occupied by the different number of channels (N) for BBO approach.

It is noted that for low value mark such as N = 4, the population size had no significant effect on the performance of BBO. From Table III it is clear that for population size of 100, the performance is significantly better as compared to other population size. But as the size of population increase the time required to get the optimized results at less iteration values slightly increase as the diversity of possible solutions increase. By carefully looking at the results, the paper fixed the population size of 30.

E. Comparison of BBO Approach with Previous Existing Algorithms in terms of Ruler Length

Table IV illustrates the total bandwidth (BW) and length of ruler (RL) occupied by different sequences obtained by a new soft computing method (BBO) for various channels 'N' and also its comparison with known OGR [24], [33], [47], [48] EQC and SA [1], [13], [21].

In literature [1] it is noted that the application of EQC and SA is limited to prime powers, so the total bandwidth and ruler length for EQC and SA are shown by a dash line in Table IV.

It is observed that the ruler length generated by BBO algorithm approaches to its optimum values that is, the results gets better. Figure 4 illustrate the comparison of BBO approach to generate optimal Golomb Ruler sequences with known OGR, EQC and SA in terms of the length of the ruler.

F. Comparison of BBO Approach with Previous Existing Algorithms in terms of Total Bandwidth

The aim to use soft computing approach (BBO) in this paper was to optimize the length of the ruler so as to conserve the total bandwidth occupied by the channels. Comparing the simulation results of BBO with known OGR, EQC and SA; it is observed that there is a significant improvement with respect to the length of the ruler (see Figure 4) and thus the total bandwidth occupied (see Table IV) by the use of soft computing methods. Figure 5 illustrate the comparison of BBO approach to generate optimal Golomb Ruler sequences with known OGR [24], [33], [47], [48] EQC and SA [1], [13], [21] in terms of the total bandwidth.

TABLE II. INFLUENCE OF INCREASE IN ITERATIONS ON TOTAL BANDWIDTH GENERATED BY SOFT COMPUTING APPROACH (BBO) FOR VARIOUS MARKS (N)

ITERATIONS	TOTAL BANDWIDTH												
	BBO												
	N=7	N=8	N=9	N=11	N=12	N=13	N=14	N=15	N=16	N=17	N=18	N=19	N=20
2	164	630	293	1003	1650	4063	5059	5861	5427	6585	16801	22228	22059
5	164	630	289	1003	1650	4063	5057	5254	5427	6585	15570	22228	22059
20	145	305	289	960	1504	3746	4569	4528	4719	6542	14362	16161	22059
50	145	238	286	672	1458	2823	3895	3889	3703	5494	10898	14714	22059
100	144	230	286	624	1286	2147	2467	3285	3647	4740	7723	13330	22059
150	144	217	286	624	1286	1979	2467	3222	3525	4541	7539	8521	22059
200	144	184	267	624	1117	1979	2293	3222	3019	4551	6187	8516	22059
500	107	168	266	610	881	1230	1803	2255	2143	3347	4449	6697	22059
1000	103	168	266	566	743	1190	1767	2188	1834	3135	3665	6331	21697
2000	84	150	259	521	683	1134	1668	1917	1834	2625	2725	5630	6106
4000	83	125	203	467	588	1049	1246	1664	1804	2239	2678	5264	5759
5000	83	125	200	440	556	1048	1177	1634	1804	2208	2566	5067	5137

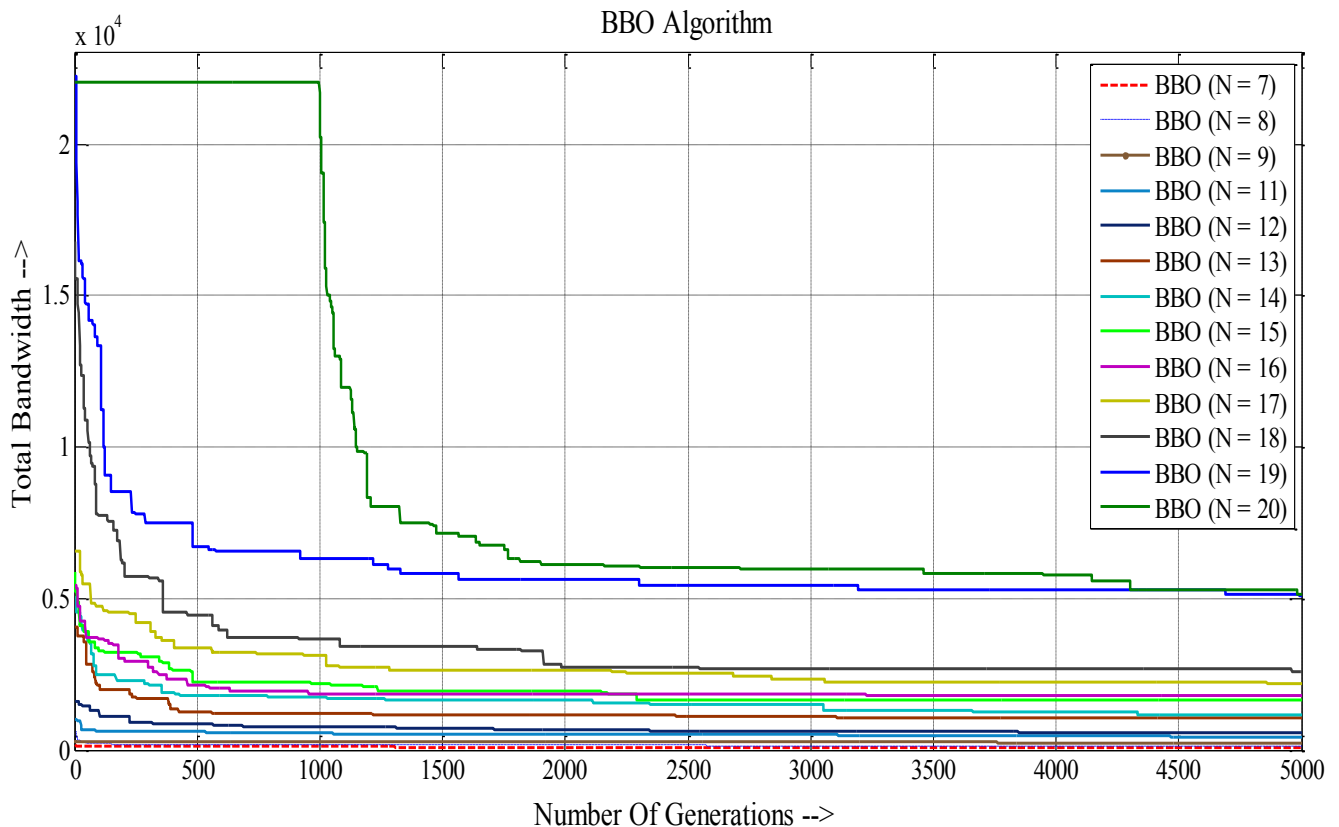


Figure 3. Influence of Generations on Total Bandwidth Obtained by BBO Algorithm for Different Values of Marks

TABLE III. INFLUENCE OF POPULATION SIZE ON THE PERFORMANCE OF SOFT COMPUTING APPROACH (BBO) FOR VARIOUS MARKS, WHERE N IS THE NUMBER OF UNEQUAL-SPACED WDM CHANNELS

POP SIZE	ITERATIONS	BBO									
		N=4		N=6		N=7		N=8		N=9	
		TOTAL BW	RL	TOTAL BW	RL	TOTAL BW	RL	TOTAL BW	RL	TOTAL BW	RL
10	5000	11	7	48	22	95	33	131	44	231	68
30	5000	11	7	42	18	83	32	131	42	206	49
50	5000	11	7	42	18	91	29	127	40	201	64
80	5000	11	6	43	20	83	32	121	39	189	69
100	5000	11	6	44	17	84	27	125	34	189	63
Here, Pop Size = Population Size, BW = Bandwidth, RL = Ruler Length											

TABLE IV. COMPARISON OF TOTAL BANDWIDTH AND RULER LENGTH OBTAINED BY SOFT COMPUTING ALGORITHM (BBO) WITH KNOWN OGR, EQC AND SA, WHERE N IS THE NUMBER OF UNEQUAL-SPACED WDM CHANNELS

N	KNOWN OGR [24], [33], [47], [48] (Best Solutions)		EQC [1], [13], [21]		SA [1], [13], [21]		BBO	
	RULER LENGTH	TOTAL BANDWIDTH	RULER LENGTH	RULER LENGTH	RULER LENGTH	TOTAL BANDWIDTH	RULER LENGTH	TOTAL BANDWIDTH
3	3	4	6	10	6	4	3	4
4	6	11	15	28	15	11	6	11
5	11	25 28	—	—	—	—	12	23
6	17	44 47 50 52	45	140	20	60	17 18 20 21	42 43 44 45 49
7	25	81 87 95 77 90	—	—	—	—	27 29 31 32 33	73 82 83 84 91 95
8	34	117	91	378	49	189	34 39 40 42	121 125 127 131
9	44	206	—	—	—	—	49 56 61 62 63 64	196 200 201 206 215 225
10	55	249	—	—	—	—	74	274
11	72	386 391	—	—	—	—	86 104 114 118	435 440 491
12	85	503	231	1441	132	682	124	556
13	106	660	—	—	—	—	203 241	1015 1048
14	127	924	325	2340	195	1183	127 206 230	924 991 1177

15	151	1047	—	—	—	—	267 298	1322 1634
16	177	1298	—	—	—	—	283	1804
17	199	1661	—	—	—	—	354 369	2201 2208
18	216	1894	561	5203	493	5100	445	2566
19	246	2225	—	—	—	—	597	5067
20	283	2794	703	7163	703	6460	752	5137

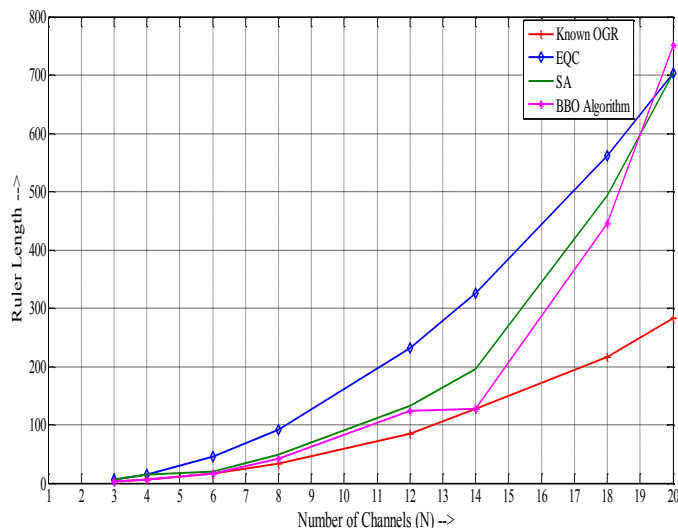


Figure 4. Comparison of the Results Obtained By BBO Approach With Known OGR, EQC and SA in Terms of Length of the Ruler

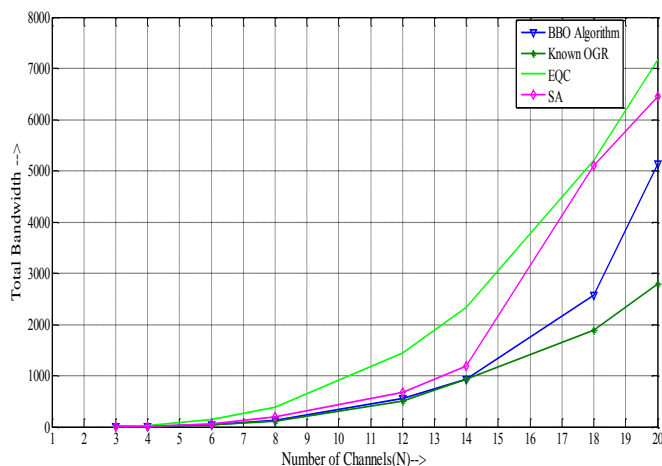


Figure 5. Comparison of BBO Approach with Known OGR, EQC and SA With Respect to the Total Bandwidth Occupied by the Ruler

VI. CONCLUSION

Generation of Optimal Golomb Ruler sequences using classical approaches is a high computational complexity problem. The purpose of soft computing is not necessarily to produce best results, but to produce the optimal results under the constraints of time and cost. This paper presents the application of BBO to solve Optimal Golomb Ruler problem. It has been observed that BBO produces Golomb Ruler sequences very efficiently. The performance is compared with the other existing classical approaches i.e. Extended Quadratic Congruence (EQC) and Search Algorithm (SA) in terms of the length of ruler and total

bandwidth obtained by the sequences. The preliminary results indicate that BBO appear to be most efficient approach to such NP-complete problems.

REFERENCES

- [1] Wing C. Kwong, and Guu-Chang Yang, "An Algebraic Approach to the Unequal-Spaced Channel-Allocation Problem in WDM Lightwave Systems", IEEE Transactions on Communications, Vol. 45, No. 3, 352–359, March 1997.
- [2] A. R. Chraplyvy, "Limitations on lightwave communications imposed by optical-fiber nonlinearities," J. Lightwave Technol., vol. 8, pp. 1548–1557, Oct. 1990.
- [3] G.P. Aggarwal, "Nonlinear Fiber Optics", Second Edition, Academic Press, San Diego, CA, 2001.
- [4] Vrizlynn L. L. Thing, P. Shum, and M. K. Rao, "Bandwidth-Efficient WDM Channel Allocation for Four-Wave Mixing-Effect Minimization", IEEE Transactions on Communications, Vol. 52, No. 12: 2184–2189, December 2004.
- [5] Nordiana Mohamad Saaid, "Nonlinear Optical Effects Suppression Methods in WDM Systems with EDFAs: A Review", International Conference on Computer and Communication Engineering (ICCCE 2010), May 2010, Kuala Lumpur, Malaysia.
- [6] Fabrizio Forghieri, R. W. Tkach, A. R. Chraplyvy, and D. Marcuse, "Reduction of Four-Wave Mixing Crosstalk in WDM Systems Using Unequally Spaced Channels" IEEE Photonics Technology Letters, Vol. 6, No. 6: 754–756, June 1994.
- [7] W. Babcock, "Intermodulation interference in radio systems", Bell Systems Technical Journal, pages: 63–73, 1953.
- [8] H. P. Sardesai, "A simple channel plan to reduce effects of nonlinearities in dense WDM systems," in Proc. Conf. Lasers, Electro-Optics, 1999, pp. 183–184.
- [9] F. Forghieri, R. W. Tkach, and A. R. Chraplyvy, "WDM systems with unequally spaced channels," J. Lightwave Technol., vol. 13, pp. 889–897, May 1995.
- [10] B. Hwang and O. K. Tonguz, "A generalized suboptimum unequally spaced channel allocation technique—Part I: In IM/DD WDM systems," IEEE Trans. Commun., vol. 46, pp. 1027–1037, Aug. 1998.
- [11] O. K. Tonguz and B. Hwang, "A generalized suboptimum unequally spaced channel allocation technique—Part II: In coherent WDM systems," IEEE Trans. Commun., vol. 46, pp. 1186–1193, Sept. 1998.
- [12] M. D. Atkinson, N. Santoro, and J. Urrutia, "Integer sets with distinct sums and differences and carrier frequency assignments for nonlinear repeaters," IEEE Trans. Commun., vol. COM-34, June 1986.
- [13] R. Randhawa, J.S. Sohal, R.S. Kaler, Optimum algorithm for WDM channel allocation for reducing four-wave mixing effects, Optik 120 (2009) 898–904.
- [14] http://www.compunity.org/events/pastevents/ewomp2004/jaillet_krajecki_pap_ew04.pdf
- [15] Gray S. Bloom and S.W. Golomb, "Applications of Numbered Undirected Graphs", Proceedings of the IEEE, Vol. 65, No. 4: 562–570, April 1977.
- [16] Vrizlynn L. L. Thing, M. K. Rao and P. Shum, "Fractional Optimal Golomb Ruler Based WDM Channel Allocation", The 8th Opto-Electronics and Communication Conference (OECC-2003), Vol. 23, pp. 631–632, October 2003.

- [17] James B. Shearer, "Some New Disjoint Golomb Rulers", IEEE Transactions on Information Theory, vol. 44, No. 7, pp. 3151–3153, Nov. 1998.
- [18] <http://theinf1.informatik.uni-jena.de/teaching/ss10/oberseminar-ss10>
- [19] Johan P. Robinson, "Optimum Golomb Rulers", IEEE Transactions on Computers, vol. C-28, No. 12, December 1979.
- [20] James B. Shearer, "Some New Optimum Golomb Rulers", IEEE Transactions on Information Theory, IT-36:183–184, January 1990.
- [21] Shobhika, "Generation of Golomb Ruler Sequences and Optimization Using Genetic Algorithm", M. E. Thesis-2005, Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Deemed University, Patiala.
- [22] Stephen W. Soliday, A. Homaifar, Gary L. Leiby, "Genetic Algorithm Approach to the Search for Golomb Rulers", Proceedings of the Sixth International Conference on Genetic Algorithms (ICGA-95), Morgan Kaufmann (1995), 528–535.
- [23] John P. Robinson, "Genetic Search for Golomb Arrays", IEEE Transactions on Information Theory, Vol. 46, No. 3: 1170–1173, May 2000.
- [24] J. Colannino, "Circular and Modular Golomb Rulers" 2003, URL: <http://cgm.cs.mcgill.ca/~athens/cs507/Projects/2003/JustinColannino/>.
- [25] A. Dimitromanolakis, "Analysis of the Golomb Ruler and the Sidon Set Problems, and Determination of Large, Near-Optimal Golomb Rulers", Master's thesis, Department of Electronic and Computer Engineering, Technical University of Crete, June 2002.
- [26] Apostolos Dollas, William T. Rankin, and David McCracken, "A New Algorithm for Golomb Ruler Derivation and Proof of the 19 Mark Ruler", IEEE Transactions on Information Theory, Vol. 44, No. 1, 379–382, January 1998.
- [27] "Project OGR", <http://www.distributed.net/OGR>.
- [28] Shonak Bansal, Shakti Kumar, Himanshu Sharma and Parvinder Bhalla, "Generation of Golomb Ruler Sequences and Optimization Using Biogeography Based Optimization", 5th International Multi Conference on Intelligent Systems, Sustainable, New and Renewable Energy Technology and Nanotechnology (IISN-2011), Institute of Science and Technology Klawad, Haryana, February 18–20, 2011, pp 282–288.
- [29] Tiago Leita, "Evolving the Maximum Segment Length of a Golomb Ruler", Genetic and Evolutionary Computation Conference, June 2004, USA.
- [30] Carlos Cotta, Ivan Dotu, Antonio J. Fernandez, and Pascal Van Hentenryck, "Local Search-Based Hybrid Algorithms for Finding Golomb Rulers", Vol. 12, Issue 3, Pages: 263 – 291, Kluwer Academic Publishers, Boston, September 2007.
- [31] William T. Rankin, "Optimal Golomb Rulers: An exhaustive parallel search implementation", M.S. thesis, Duke University, 1993. (Available at <http://people.ee.duke.edu/~wrankin/golomb/golomb.html>).
- [32] <http://mathworld.wolfram.com/PerfectRuler.html>
- [33] <http://mathworld.wolfram.com/GolombRuler.html>
- [34] Hongwei Mo, Lifang Xu, "Biogeography based Optimization for Travelling Salesman Problem", Sixth International Conference on Natural Computation (ICNC 2010), vol. 6, pp.3143–3147, August 2010.
- [35] D. Simon, M. Ergezer, D. Du, and R. Rarick, "Markov models for biogeography-based optimization", IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 41, no. 1, pp. 299–306, January 2011.
- [36] Shakti Kumar, Parvinder Bhalla and Amarpartap Singh, "Fuzzy Rule Base Generation Part II: A BBO Approach", Proceedings of the Workshop on Intelligent System Engineering (WISE-2010), Institute of Science & Technology, Klawad, July 2010.
- [37] S. Kumar, P. Narula, and T. Ahmed, "Knowledge extraction from numerical data for the Mamdani type fuzzy systems: A BBO approach", Innovative Practices in Management and Information Technology for Excellence, Jagadhri, India, May 2010.
- [38] H. Ma, S. Ni, and M. Sun, "Equilibrium species counts and migration model tradeoffs for biogeography-based optimization", IEEE Conference on Decision and Control, Shanghai, pp. 3306–3310, December 2009.
- [39] H. Ma, "An analysis of the equilibrium of migration models for biogeography-based optimization", Information Sciences, vol. 180, no. 18, pp. 3444–3464, 15 September 2010.
- [40] A. Wallace, "The Geographical Distribution of Animals (Two Volumes)", Boston, MA: Adamant Media Corporation, 2005.
- [41] C. Darwin, "The Origin of Species", New York: Gramercy, 1995.
- [42] Dan Simon, Mehmet Ergezer, and Dawei Du, "Population Distributions in Biogeography-Based Optimization Algorithms with Elitism", Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA, pp. 991–996, October 2009.
- [43] D. Simon, "Biogeography-Based Optimization," IEEE Trans. on Evolutionary Computation, vol. 12, no. 6, pp. 702–713, December 2008.
- [44] Aniruddha Bhattacharya and Pranab Kumar Chattopadhyay, "Hybrid Differential Evolution With Biogeography-Based Optimization for Solution of Economic Load Dispatch", IEEE Transactions on Power Systems, vol. 25, no. 4, pp. 1955–1964, November 2010.
- [45] S. Kumar, P. Bhalla, and A. Singh, Fuzzy rule base generation from numerical data using biogeography-based optimization, Institution of Engineers Journal of Electronics and Telecomm Engineering, vol. 90, pp. 8–13, July 2009.
- [46] H. Ma and D. Simon, "Blended biogeography-based optimization for constrained optimization", Engineering Applications of Artificial Intelligence, vol. 24, no. 3, pp. 517–525, April 2011.
- [47] <http://www.research.ibm.com/people/s/shearer/grtab.html>
- [48] <http://www.research.ibm.com/people/s/shearer/gropt.html>
- [49] Rudra Pratap, "Getting Started with Matlab A Quick Introduction for Scientists and Engineers", Oxford University Press, New York, 2010.

APPENDIX – A

The table below shows the optimal Golomb Ruler (OGR) sequences generated by Biogeography Based Optimization (BBO) for various marks:

TABLE V. OPTIMAL GOLOMB RULER SEQUENCES GENERATED BY BBO ALGORITHM

Order	Length	Marks
1	0	0
2	1	0 1
3	3	0 1 3
4	6 7	0 1 4 6 0 1 3 7
5	12	0 1 3 7 12
6	17 18 18 20 20 21 21 22	0 1 4 10 12 17 0 1 3 8 12 18 1 2 4 9 13 19 1 2 4 8 13 21 0 1 3 7 12 20 0 1 4 6 13 21 1 2 5 7 14 22 0 2 5 6 13 22
7	27 29 29 31 31 31 31 31 32 32 32 33	1 2 4 9 18 22 28 1 3 6 12 13 26 30 2 6 8 9 18 23 31 0 1 3 7 18 23 31 1 2 5 11 13 18 32 0 1 3 8 12 18 31 1 2 4 9 13 19 32 0 1 4 9 15 22 32 2 5 9 10 19 21 34 2 3 5 9 18 23 35
8	34 39 40 42	1 2 5 10 16 23 33 35 1 2 4 9 15 19 31 40 1 2 5 12 18 20 32 41 1 2 8 10 13 23 27 43
9	49 56 61 62 63 64	1 5 11 12 20 33 36 38 50 0 1 5 8 19 25 35 47 56 0 4 5 7 17 23 31 52 61 1 3 6 7 16 23 44 52 63 1 2 6 12 14 34 37 55 64 0 2 5 12 13 27 31 47 64
10	74	0 3 5 13 22 28 29 40 60 74
11	86 104 114 118	0 4 12 18 25 28 55 60 75 77 86 9 14 20 23 27 35 54 76 77 93 113 3 4 9 13 21 28 49 51 62 78 117 3 4 8 15 18 37 53 55 80 97 121
12	138	2 3 9 13 18 21 43 57 70 94 120 140
13	203	1 9 14 29 40 41 63 70 123 141 147 166 204
14	127 169 206 230	0 5 28 38 41 49 50 68 75 92 107 121 123 127 0 7 15 24 34 45 57 70 84 99 115 132 150 169 2 3 5 9 17 30 50 67 86 96 126 135 157 208 3 5 13 16 35 52 58 79 95 104 130 135 219 233
15	267 298	1 3 28 32 38 43 46 62 90 111 131 143 144 182 268 7 9 10 19 41 59 70 76 103 124 140 179 225 267 305
16	283	3 4 7 17 36 56 79 81 87 125 142 166 192 258 265 286
17	354 369	0 2 7 15 21 62 66 90 99 116 138 169 172 243 311 343 354 2 5 6 14 21 32 49 54 108 110 180 190 222 247 253 337 371
18	445	0 1 3 17 29 35 71 98 102 122 147 160 212 235 256 295 338 445
19	597	9 21 76 80 91 120 188 207 224 227 272 303 396 401 443 457 465 481 606

20	752	0 20 24 56 73 81 118 136 176 188 202 207 218 372 381 455 483 531 664 752
----	-----	---

Improving Enterprise Access Security Using RFID

Dr. zakaria Saleh, Yarmouk
University
Irbid, Jordan
zzaatreh@yu.edu.jo,

Dr Izzat Alsmadi, Yarmouk
University
Irbid, Jordan
ialsmadi@yu.edu.jo,

Ahmed Mashhour Yarmouk
University Irbid, Jordan
mashhour@yu.edu.jo

Abstract—Personal Computers now a day are widely used as workstations on many organizations networks. Hence, the securities of the workstations become an integral part of the overall security of the network. Consequently, any good access control solution should be designed in such a manner that key information cannot be retrieved without proper authentication. RFID can be used as an alternative for providing extended user authentication. This study believes that the most secure methods include storing the access information on another secure device such as a smart card, or an RFID tag. Standard operations require that workstation to be configured in a way that involves interactive user authentication is instead of an automatic login where the password is stored on the workstation. Using an RFID system will insure that this requirement is kept intact. Many security systems fail not because of technical reasons, but because of the people who could protect a system were not following the basic security standards like locking the workstation before moving away. The proposed RFID system will enforce locking the workstation as soon as the user moves away from that computer unit.

Keywords: RFID, Workstation Security, Authentication, Access Managers

I. INTRODUCTION

All computer systems contain vulnerabilities, and one of the most significant vulnerabilities is the user (intentionally or accidentally). The best way to protect a workstation and the confidentiality of data it holds, is when access control is implemented, the access control should be hardware based so that the control is maintained as soon as possible in the during system startup and access. In addition, when a user wants to leave the workstation unattended for a period of time without powering off, sound security practice requires that no unauthorized access is allowed to the system in the user's absence. This paper will concentrate on user authentication and prevention of (or protection against) access to work station by unauthorized user, and ensuring that users are the persons they claim to be with the ability to

protect information and system resources. System resources include CPUs, disks, and programs, in addition to information on the work station. Classically, access control logon sequences have required a user name and password combination to verify the identity of a user. This research will introduce biometric devices capable of reliably identifying users through an RFID system.

II. SIGNIFICANCE OF THE STUDY

All computer systems contain vulnerabilities, and one of the most significant vulnerabilities is the user [6]. Anytime a workstation is running and not locked, the workstation can be vulnerable and convenient to be used by an unauthorized person in the work place. Thus, user authentication is a required component of all workstations, not only at startup or log on, but while the system is being used as well to protect information assets from deliberate or unintentional unauthorized acquisition, disclosure, manipulation, modification, damage, loss, or use. Many security systems fail not so much for technical reasons, often the people who could protect a system were not the ones who suffered the costs of failure [7]. User authentication is the backbone of any access control solution. Therefore, it is important that any good workstation security measure should provide a very high integrity user authentication solution. The proposed security enhancement of using RFID as an authentication means with continuance monitoring of the RFID tag, used to run the workstation, will insure a secure system that is impossible for unauthorized persons to break into. The RFID tag has adequate secure storage to store access control profiles. The major disadvantage of a using RFID is the necessity for supplying a An RFID reading device on each workstation. However, with the current price for RFID readers, this may be justified.

III. WORKSTATION SECURITY OVERVIEW

Security is the process of preventing unauthorized use of a computer or a workstation. The traditional foundation of

workstation security is based on implementing safeguards to ensure that users access only the resources and services that they are entitled to access. In addition, measures are taken so that qualified users are not denied access to services that they are expecting to receive. Absolute prevention is theoretical, and If a computer is compromised, the entire contents of the system are exposed to the attacker[6].

For any workstation, authentication can be done by one of three ways 4: Something the user knows (e.g., a password); something the user has (e.g., a token or card); something the user is (e.g., fingerprint, voice, eye scan). Each approach has advantages, and limitations. This paper is more concern with the limitation part:

1. "Something the user knows" can be forgotten, guessed by others, or inappropriately shared,
2. "Something the user has" can be misplaced or stolen, and
3. "Something the user is" can be difficult to distinguish reliably.

Therefore, combining two or more methods enhances the confidence level (e.g. a bank ATM machine requires both a card and a password). However, while an access control system must be effective, it should also be user friendly [1].

Currently, Windows and workstation authentication uses or depends on the first type of authentication techniques. Mixing this with RFID authentication (i.e. something the user has), will improve security and reduce the possible of wrongly indentifying a user.

When a user logs on to a computer running Microsoft Windows for example, the user needs to supply a user name and password. This becomes the default security context for connecting to other computers on networks and over the Internet. Thus, passwords are an important aspect of computer security. They are the front line of protection for user accounts. A poorly chosen password may result in the compromise of the entire corporate network. Passwords are still the most pervasive tool used to secure access to networks and databases. As the number of passwords per employee increases, the likelihood of them being forgotten rises [2]. For maximum security each member required to protect their password. Access can further protected by following good password practices (e.g. creating passwords that are a mix of letters, numbers, and other characters). Depending on the level of security needed, users can choose from standard to very high levels of password security.

A security breach in accessibility occurs when either access for a system is denied for an authorized user or access (an example of this category would be an authorized user of a system who is unable to access a system due to forgetting their password)[3]. To make passwords that are easy to remember, many people create passwords that contain their name or email address, or are a string of familiar digits, such as their phone number or birthday. The problem is, simple

passwords like this are easy for intruders to guess, and could compromise the security of the network. Users accessing highly sensitive data on the network, need to employ "complex" passwords (e.g. passwords that do not contain parts of users name or birthday are complex), however, extensive password requirements can overload human memory capabilities as the number of passwords and their complexity level increases [3].

IV. ACCESS OR ACCOUNT MANAGERS

In Web application security deployments, and many other types of distributed systems, users accessing a protected application are authenticated via enterprise identity/access management products, such as Netegrity's SiteMinder, IBM's WebSEAL, and Oracle access manager. The authorization service, however, is delegated to the provider of the application itself, or to the application server. Generally, there are major goals or requirements for any access or account manager. Those are:

- Provide a single username and password.
- Accept alternative forms of authentication (such as RFID) beyond username/password
- Provide strong authentication mechanisms where needed
- Provide single sign on (SSO) where possible.
- Provide strong security that does not slow performance.

Most access managers provide an authentication API for integrating a variety of authentication methods and devices such as smart cards. Account manager information are usually updated to stay in synchronization with account in LDAP or active directory.

V. AUTHENTICATION

Most current access managers are designed to deal with different types of authentication. This may include: Basic username/password, X.509 Certificates, Smart Cards, Two factor tokens, Form-based, and Custom authentications via Authentication APIs.

VI. LDAP

Lightweight directory access protocol (LDAP) is a directory service protocol that provides access to a directory over a network. It stores information in directory service (such as Microsoft Active Directory) and query it.

VII. RELATED WORK

There are several applications related to using RFIDs in security and authentication [5], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], and [21]. This paper followed the trend of the majority of the papers that are discussing RFID where they present using RFID for a particular application. This may span from generic applications that can be applied in several domains such as users' authentication (e.g. students, employees, citizens, etc). In such applications, RFID authentication is used as an

alternative, more convenient authentication service for some other typical authentication tools such as biometrics, software authentication, etc. In general, authentication methods can be classified into 3 categories for users: something they are (e.g. biometrics, such as fingerprints, voice, etc), something they say, know or type such as passwords, and something they have such as the physical keys and the access or RFID cards. For better security, many entities are trying to combine methods from the different categories.

The second type of papers talking about RFID discusses security concerns and issues in the RFID network itself. Examples of such papers that discussed security and vulnerability issues in RFID networks are [5], [12], [14], [15], [18], and [21].

Ham et al studied merging RFID with PKI and DNS security extensions for establishing a secure network [8]. The DNS with security extensions can provide integrity and data authentication. Mao et al proposed an Interoperable Internet-Scale Security (IISS) framework for RFID networks on which multiple partners with different identity schemes can be authenticated [9]. The framework made authentications based on an aggregation of business context, enterprise information, and RFID tag information as a lightweight solution for the problem of relations trust authentication in RFID networks.

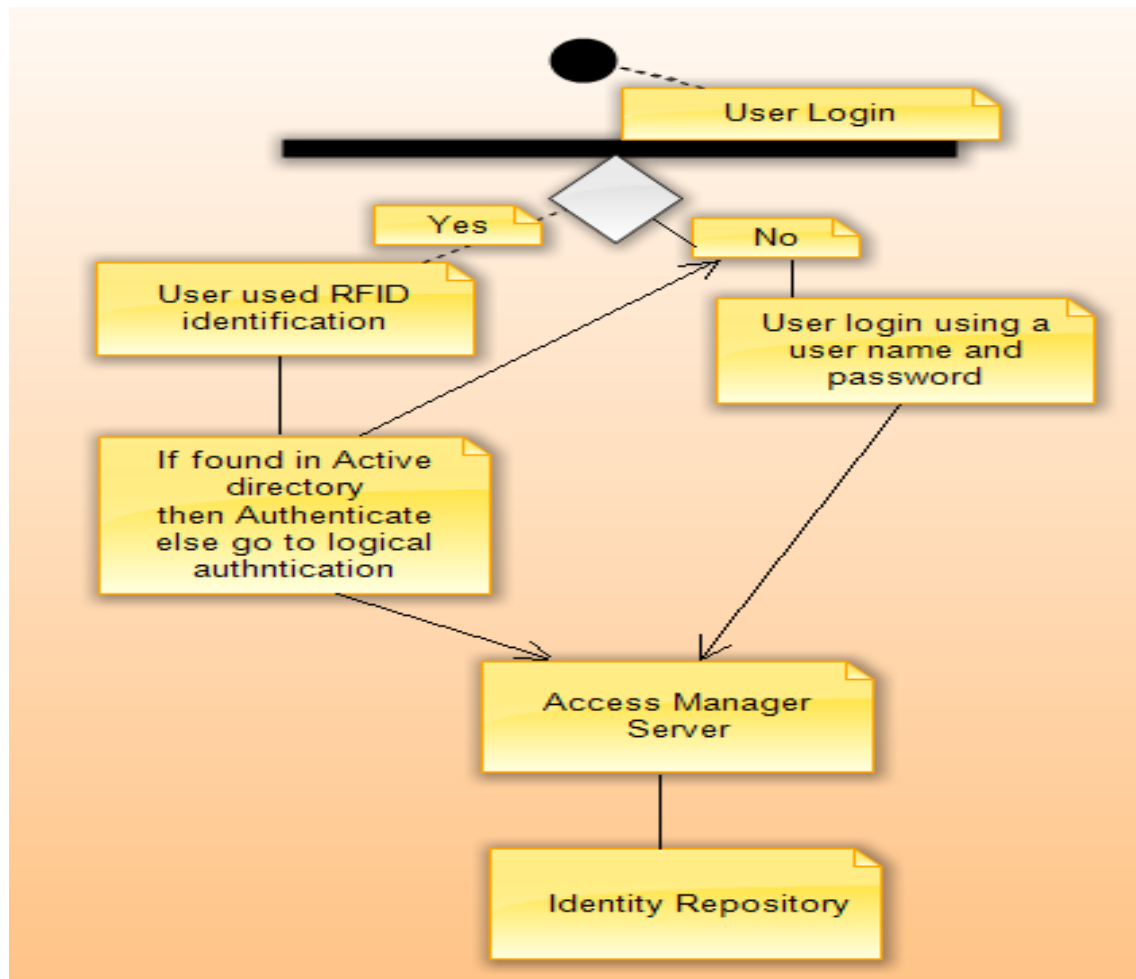


Figure1. Proposed modification on authentication systems to include RFID authentication.

Zhao et al proposed a hierarchical P2P based RFID code resolution network structure In order to alleviate or solve some performance and security problems of RFID code resolution [10]. RFID code resolution services and related security mechanism are implemented. Ku et al presents a complex event mining network which enables automatic and

real-time routing, caching, filtering, aggregation and processing of RFID events and defines the fundamentals of RFID enabled supply chain event management [11]. Kim et al propose the modified hash based RFID security protocol to improve data privacy and authentication between a tag and a

reader [12]. The paper discussed some of the vulnerabilities that may occur in the RFID network.

Chang et al proposed a method similar to the one adopted in this research in combining RFID with cell phones for users' authentication [13]. They also studied security and vulnerability issues in RFID networks. To achieve message security, it is essential to keep anonymity to protect the privacy of the RFID credit card holders.

VIII. DESIGN AND APPROACHES

Figure 1 shows a simplified diagram for the proposed modification on workstations authentication system. RFID cards can be connected to the workstation through wireless that enable users to be granted login once they are close enough (in a defined distance that depends on power and frequency). In order to simply system recognizing users and correlate users with RFIDs, RFID values can be generated using a seed value correlated with the user information. Proposed modification should guarantee Single Sign On (SSO) where user will be asked only once to verify their identity. Once system found a possible problem in authentication, it may ask for the second type of

IX. RFID RANGE AND FREQUENCY

Selecting the proper frequency for this RFID is significant. Recommended Frequency is 13.56 MHz. This frequency has several characteristics that may make it suitable. This include: low cost, ultra-thin, battery-less contactless read/write technology (approximate read range up to 1.5 meter), and offers increased and advanced security over 125 KHz proximity systems. The technology is capable of providing advanced security features like encryption algorithms, where each transponder has a unique tamper proof factory programmed ID code.

The RFID range selection is fundamental. If you're planning to use RFID you need to know what distance it will work over. For a computer workstation or server in a room, the typical distance that those equipments exist in may vary between 2 – 30 square meters. Besides frequency, there are several other parameters that regulate the RFID transmitting and receiving distance. Those other parameters include: RF transmit power, the receive sensitivity, the surroundings, how much water is present, the orientation of the tag, and the care that's gone into designing the products, planning and installing the system. Liquids such as water can absorb RF (especially at microwave frequencies) and metals can shield or reflect RF energy.

In terms of the power, the RFID component attached to the computer should not have a problem as it can be simply a USB extension which can take power through the USB port. For simplicity, the RFID part that will be attached to the employee card can be a simple active RFID tag can receive its power from a small battery or passive tags that can get their power from the RFID transmitter attached to the computer. Currently several companies such as Noxel (www.noxel.com/rfid-reader.html) and Gemia are developing RFID readers

authentication. Users will be logged of whenever they leave the close distance range defined.

The proposed modification on authentication assuming that users' machines will be locked as soon as they leave them. Many users avoid locking screens as it is inconvenient for them to lock the screen and type passwords again and again over the day. As such, a solution is to have a program that automatically detect the user RFID whenever the user comes close to the machine. This can be very simple through implementing transceivers between the computers and the RFID. In most cases, however, we may need only one way communication where the RFID will transmit their ID to desktops.

The transmitted signal should be modulated or encrypted with the user information for two reasons: First, this is to guarantee that signals will not be intercepted in the middle and saved and possibly reused by intruders. On the other hand, this is a double identification matching technique where each RFID unique number will be attached to a particular user in which there is always a one to one relation between users' and their RFID.

using Bluetooth technologies to combine those two technologies and eliminate the need to connect the RFID reader with the computer through a wire.

X. EXPERIMENT AND EVALUATION

In order to demonstrate the approach, we implemented the system and develop a program with RFID using USB connection. Such test can validate many features of the proposed system except those related to the required distance between the computer and the user for the program to detect the RFID and some other issues possibly related to security.

In the developed program, the program is started as a service and always in listening or receiving state, similar to those happened in socket programming such as chat or messaging services. As soon as users enter the RFID card in the reader, the RFID information are sent to the LDAP to verify the user identity using the information saved in the LDAP or the active directory about users that include user relevant RFID. This information should be encrypted and read only by system applications similar to passwords.

XI. UNIVERSITY CAMPUS, A CASE STUDY

In order to assess the design and specification requirements for an RFID system, a small subset of Yarmouk University campus is selected. This represents the IT faculty which comprises of two major building with an approximate distance of 20-30 meters between those two building. An RFID simulator (Turck Inc.). Number of users based on computer workstations and servers is approximated to be 100 computer and server. This excludes computers in the labs as those computers are usually public and should not include private logins. Besides the number of RFID elements, the major

attributes selected in the simulation are distance, speed (of message transmission) and data quantity. Those 3 elements are adjustable in the simulator as they impact each other and the overall simulation process.

Read/Write distance is set at the range of: 0-40. While data quantity is not expected to be a major issue in the access verification scenario where the amount of data to transfer is minimal (i.e. that is required for authentication). This is different from other scenarios such as warehouse or store management where it is expected to have a large amount of data transmission among RFID system components. Nonetheless, speed is important and the speed of response by the simulators is set to the minimum to ensure that the logging system will not be a bottleneck and affect the overall working environment.

XII. CONCLUSION AND FUTURE WORK

In this paper, we proposed using RFID to improve enterprise access security through combining typical software or logical security with RFID. This combination is expected to improve the overall security infrastructure of distributed systems while at this same do not impact the system performance or causing extra overhead elements.

RFID security access control system can be added to the existed infrastructure without the need for significant extra software or hardware elements. An elementary simulation is implemented to demonstrate the proposal and evaluate the major elements that can impact selecting the RFID security such as data quantity, speed and distance. Results showed that such security infrastructure can be applicable for local area distributed system as such University campuses, schools, warehouses, and small to medium size enterprises.

REFERENCE

- [1] Graham, I (1996). "PC Workstation Security" A paper presented by 1996 Information Security Summit on 29-31 May, 1996 at the Tattersal's Club, Sydney.
- [2] Bjorn, V. (2006)"Solving the Weakest Link in Financial Institutions Network Security: Passwords". A Digital Persona, Inc. White Paper, September 2006.
- [3] Carstens, D. & McCauley-Bell, P.(2004). "Evaluation of the Human Impact of Password Authentication Practices on Information Security". Informing Science Journal, Vol 7, 2004.
- [4] Kolodgy, C. (2001). "Biometrics: You Are Your Own Key". InfoWorld (January 29, 2001) Issue.
- [5] Park, N., Choi, D., Kim, S., and Won, D. (2008). Enforcing Security in Mobile RFID Networks Multilateral Approaches and Solutions, IEEE.
- [6] PNNL (2010). "2010 Guide for Home Computer Security". Pacific Northwest National Laboratory. Retrieved for the WWW on April 6, 2010 from www.pnl.gov/media/homeguide_public.pdf.
- [7] ANDERSON, R., & SCHNEIER, B. (2005). "Economics of Information Security". IEEE COMPUTER SOCIETY, vol. 3 no. 1.
- [8] A Study on Establishment of Secure RFID, Network Using DNS Security Extension, YoungHwan Ham *, NaeSoo Kim *, CheolSig Pyo*, JinWook Chung, 2005 Asia-Pacific Conference on Communications, Perth, Western Australia, 3 - 5 October 2005.
- [9] An Interoperable Internet Scale Solution for RFID Network Security, Tingting Mao, John R. Williams, Abel Sanchez, 2009 IEEE.
- [10] Research on hierarchical P2P based RFID code resolution network and its security, Wen Zhao, Xueyang Liu, Xinpeng Li, Dianxing Liu, Shikun Zhang, 2009 International Conference on Frontier of Computer Science and Technology.
- [11] Novel Complex Event Mining Network for RFID-Enable Supply Chain Information Security, Tao Ku1, 2 YunLong Zhu1 KunYuan Hu1, 2008 International Conference on Computational Intelligence and Security.
- [12] Analysis of the RFID Security Protocol for Secure Smart Home Network, Hyun-Seok Kim, Jung-Hyun Oh, and Jin-Young Choi, 2006 International Conference on Hybrid Information Technology (ICHIT'06)
- [13] An Improved Certificate Mechanism for Transactions Using Radio Frequency Identification Enabled Mobile Phone, Allen Y. Chang, Dwen-Ren Tsai , Chang-Lung Tsai , Yong-Jiang Lin , 2009 IEEE
- [14] Intrusion Detection in RFID Systems, Geethapriya Thamilarasu and Ramalingam Sridhar, 2008 IEEE
- [15] Trust and Security in RFID-Based Product Authentication Systems, Mikko O. Lehtonen, Member, IEEE, Florian Michahelles, and Elgar Fleisch, IEEE SYSTEMS JOURNAL, VOL. 1, NO. 2, DECEMBER 2007.
- [16] A Layered Approach to Design of Light-Weight Middleware Systems for Mobile RFID Security, (SMRM : Secure Mobile RFID Middleware System), Namje Park, Jooyoung Lee, Howon Kim, Kyoil Chung, and Sungwon Sohn,
- [17] Engineering Management-Focused Radio Frequency Identification (RFID) Model Solutions, —PAUL G. RANKY, IEEE ENGINEERING MANAGEMENT REVIEW, VOL. 35, NO. 2, SECOND QUARTER 2007.
- [18] The RFID Middleware System Supporting Context-Aware Access Control Service, Jieun Song and Howon Kim, Feb..20-22, 2006 ICAOT2006.
- [19] NOVEL RFID-BASED SHIPPING CONTAINERS LOCATION AND IDENTIFICATION SOLUTION IN MULTIMODAL TRANSPORT, Zhengwu Yuan, Dongli Huang, CCECE/CCGEI May 5-7 2008 Niagara Falls, Canada.
- [20] RFID for airport security and efficiency, Thomas McCoy, R Bullock and P Brennan, IEE.
- [21] Secure and Efficient Recommendation Service of RFID System using Authenticated Key Management, Jinsu Kim1, Changwoo Song, Taeyong Kim, Keewook Rim, Junghyun Lee, 2009, IEEE.

AUTHORS PROFILE

Zakaria Saleh: Dr. Zakaria Saleh is an associate professor in the Faculty of IT and Computer Sciences, at Yarmouk University. His work experience ranged for simply providing technical support and nonconformance resolutions for a "Compaq Computers" PC configuration center, to working on the design and development of electronic control systems in the Automotive Industry,

where he has contributed to the introduction of M2M (Machine to Machine) Communication Systems. Prior to joining Yarmouk's Faculty Team, he was working as a Project Engineer, at Case Corporation, an International Designer and Manufacturer of Agricultural and Construction Equipment, located in the USA. He was a member of the engineering team where he has contributed to the design and

development of several microcontrollers, and was the lead engineer to work on the design and development of web based Fleet Management System.

Izzat Alsmadi: Dr Izzat Mahmoud Alsmadi is an assistant professor in the department of computer information systems at Yarmouk University in Jordan. He obtained his Ph.D degree in software engineering from NDSU (USA), his second master in software engineering from NDSU (USA) and his first master degree in CIS from University of Phoenix (USA). He has a B.sc degree in telecommunication engineering from Mutah university in Jordan. Before joining Yarmouk University he worked for several years in several

computer science and information technology companies and institutions in Jordan, USA and UAE.

His research interests include: software engineering, software testing, software metrics and formal methods.

Ahmad Mashhour: Dr. Ahmad Mashhour earned his PhD degree from the University of London (LSE) 1989 in Information Systems. He is currently a faculty member at Yarmouk University, Jordan. He worked as a visiting professor at University of Qatar, and then at the University of Bahrain. His current research interest includes information systems modeling and analysis, information systems security, e-Business, and e-learning.

Enhancement of stakeholders participations in Water fall Process Model

(Step towards reducing the defects in software product)

Mehar Ullah
meharullah@yahoo.com

Fasee Ullah
faseekhan@gmail.com

Muhammad Saeed Shehzad
saeedshehzad@gmail.com

Department of Computer Science
City University of Science & Information Technology (CUSIT)
Peshawar, Pakistan

Abstract— In complete software development life cycle, defects can be originated from any source such as from stakeholders, end users, or in understanding the scope of project or product. In water fall process model, all activities are performing in sequence and though it has its own drawbacks, which causes of different defects but one perspective of defects is the involvement of developer stakeholders in development process. The coordination problem between developer stakeholders of successive activities causes many problems such as improving defect ratios, managing the work within deadline time, productivity, reliability and quality of software. Coordination and communication problem among stakeholders is due to lack of communication power of stakeholders and proper way to understand his/her work to stakeholder of successive activity. To overcome this problem, we have proposed a strategy which can be implemented by project manager of team or with mutual coordination of team members.

Key Words: Defects , stakeholders, Defects ratio, Coordination, Communication

I. INTRODUCTION

Software development process comprises on set of activities which can be shaped or named according to define methodologies and umbrella of these activities is considered as process model. Now-a-day, stakeholders are using number of process model and their demand can be seen with respect to different aspect such as delivery time for products, quality level, maintainability, availability, complexity or agility. Among these process model, water fall is an old and traditional model which can be followed by many developers to develop the customized software and where instant change in system are not acceptable. Water fall process model is also represented as classic software life cycle [1] where software evolution proceeds in sequence of activities. Besides its advantages, water fall process model causes some problems due to its sequential approach, making the development process length and unable to accept the uncertain requirements of a system [2]. Similarly, in waterfall process model planning

is done during the early stages, so it causes many design flaws before the development process. But its planning and intensive documentation helps to maintain the product quality. For considering the full waterfall process model, developers can use the set of activities such as system requirement, software requirement, architectural design, detail design, coding, testing and maintenance [3, 4].

In each stage of waterfall process model, documents are created to describe the objectives and requirements of that phase and at the end of each phase a review of project is held for continuation on next phase [5, 6, 7]. But if developer stakeholder of current stage is unable to communicate effectively with developer stakeholder of next phase then number of factors arise which can impact the achieving of functional or non functional requirements, delay in delivery of product and its defect rate. Similarly, external influence of software development causes the risk factor which can lead further to cost, duration and quality of projects [8].

In 1960, some software crises come in front of audience during development phase. Later on in 1993, an IEEE standard defines several dimensions of defects that should be collected [9]. There are number of interrelated factors in documentation, process management, development and activities sequences which cause defects but most probably communication gap between stakeholders of successive phase is considered as important source [10, 11].

To overcome this problem, we have proposed a strategy to fill the communication gap between stakeholders of two connective phases and reduce the defect rate.

II. PROPOSED METHODOLOGY

In water fall model, development of software is done by following a set of activities in sequence and each activity is performed by one or more than one stakeholders. The

coordination problem between developer stakeholders of successive activities causes many problems such as improving defect ratios, managing the work within deadline time, productivity, reliability and quality of software. Coordination

and communication problem among stakeholders is due to lack of communication power of stakeholder and proper way to understand the work of stakeholder of successive activity. To overcome this problem, a proposed strategy which can be implemented by project manager of team or with mutual coordination of team members. According to this proposed strategy the work of each stakeholder should be documented for easy access and help to stakeholder(s) of next phase. Influence of proposed work over the activities of water fall model is shown in Figure 1.

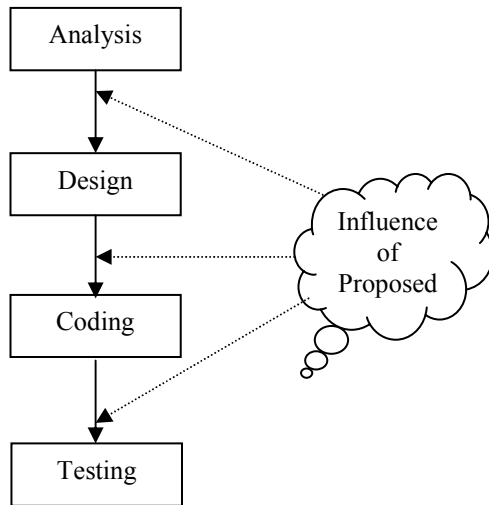


Figure 1. Influence of proposed strategy and activities list of Water fall Model

Figure 1 shows the list of activities and implementation of proposed methodology whose influence will be remain during the phase, but here it has shown only at the end of activity or phase. In proposed methodology, we considered the five roles who worked together under supervision of a project manager. These roles are of project manager, system analyst, designer, programmer and tester. Each role will follow the rules which are defined in methodology. But here in next section only the rules and work of system analyst according to methodology is defined.

The first activity of waterfall model is the analysis or requirement specification and in this phase main role is of system analyst. Instead of his/her actual work, each analyst will must use the following sequence shown in Figure 2.

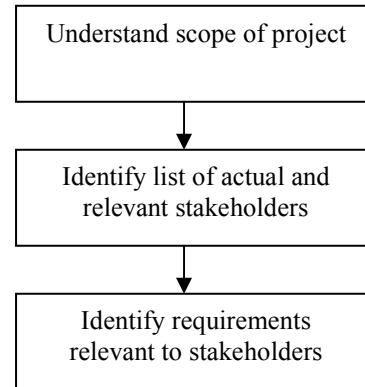


Figure 2. Steps for System Analyst in proposed methodology

Information about users, their requirements and some other information is shown in table 1.

Table 1, show the information which is maintained by system analyst for precise communication with stakeholder(s) of next coming phase or activity. The first column of table 1 shows the list of all linked and non-linked departments from where requirements are collected. The second column represents the list of users who are involved in operations directly or indirectly. Third and forth columns represent the management level of users and their assign roles respectively. Moreover, fifth column show the list of requirements which are gathered from different users of proposed system. Finally, last column represent the page number of feasibility report where gathered requirements have been organized.

III. CONCLUSION AND FUTURE WORK

In water fall process model, communication gap and understanding between developer stakeholders of successive stages causes of many defects and its effect on the maintenance period of product. Because, due to maintenance process extra efforts are needed to overcome the problems and reducing the defect rate. Due to proposed methodology in this paper, developer becomes able to convey their messages and enhance the understandability of his/her work to the stakeholder of next coming stage or phase. Here, author has presented the rules and task for system analyst only and this thing has been defined for other type of developer stakeholders. Finally, author's proposed strategy can be enhanced and precise after its implementation for customized projects and according to opinion of developer stakeholders.

Department Name	User Name	Level	Working as (role)	Requirement(s)	Page# of feasibility report to show the organizing of requirement
Admission	U1	Low	KPO	R1	N
				R2	N

	U2	Middle	Admission Officer	Rn	N
				R1	N
				R2	N

				Rn	N

Table 1: Precise information for designer

REFERENCES

- [1] Walt Scacchi, "Process Models in Software Engineering", J.J. Marciniak (ed.), Encyclopedia of Software Engineering, Feb 2010.
- [2] Center for Technology in Government University at Albany / SUNY, "A Survey of System Development Process Models", Center for Technology in Government University at Albany / SUNY, 1998.
- [3] Nabil Mohammed Ali Munassar and A. Govardhan, A Comparison Between Five Models Of Software Engineering, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010.
- [4] PK. RAGUNATH ET AL, Evolving A New Model (SDLC Model-2010) For Software Development Life Cycle (SDLC), IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.1, January 2010.
- [5] Craig Larman, "Iterative and Incremental Development: A Brief History", Published by the IEEE Computer Society, July 2003.
- [6] <http://www.buzzle.com/articles/waterfall-model-advantages-and-disadvantages.html>
- [7] IEEE Std 1044-1993. IEEE Standard Classification for Software Anomalies, 1993.
- [8] Raymond Madachy, Barry Boehm and Dan Houston, "Modeling Software Defect Dynamics", STN 13-1 April 2010.
- [9] Sakthi Kumaresh and R Baskaran, "Analysis and Prevention for Software Process Quality Improvement", International Journal of Computer Applications (0975 – 8887) Volume 8– No.7, October 2010.
- [10] Raymond Madachy, Barry Boehm and Dan Houston, "Modeling Software Defect Dynamics", STN 13-1 April 2010.

AUTHORS PROFILE

Mr. Mehar Ullah is a Lecturer in Computer Science Department Kardan Institute of Higher Education, Kabul Afghanistan. The author pursuing his MS(Software Engineering) from City University of Science & IT.

Mr. Fasee Ullah is a lecturer and active researcher in the field of Networks and System Security. He is currently working at Department of Computer Science, City University of Science & IT. He has done MS (IT) from SZABIST, Pakistan and currently is official reviewer of IEEE committee.

Mr. Muhammad Saeed Shehzad is working as Assistant Professor in the department of computer science department, City University of Science and Information Technology. He has done his MS in software engineering from City University of Science & Information Technology – Pakistan.

Loopholes in Secure Socket layer and Sniffing

Amit Mishra

Department of Computer Science & Engg., Faculty of Engineering & Technology
Jodhpur National University
Jodhpur, India
mishranamit2211@gmail.com

Abstract— Network sniffing was considered as a major threat to network and web application. Every device connected to the Ethernet-network receives all the data that is passed on the segment. By default the network card processes only data that is addressed to it. However listening programs turn network card in a mode of reception of all packets – called promiscuous mode. So, a *sniffer* is a special program or piece of code that put the Network Interface Card (NIC) in the promiscuous mode. When NIC works in promiscuous mode, the user of that system can steal all the data including password etc. without generating any traffic. Any network system running the sniffer can see all the data movement over the network. Many sniffers like Wireshark, Cain & Abel, ethersniff etc. are available at no cost on the internet. There are many proposed solutions are available for the detection of network sniffing including antisniff [1], SnifferWall [2], Sniffer Detector [3] etc. but any solution does not guarantee full security. Due to this reason many new techniques were developed including secure socket layer (https), one time password etc. but now there are some techniques that can be used to sniff this secure data. In this paper we are discussing different aspects of sniffing, methods to sniff data over secure socket network and detection of sniffer. The paper describes all the technical details and methods to perform this task.

Keywords- network sniffer; ethernet; LAN; ARP; SSH; ping

I. INTRODUCTION

Computer networks are the backbone of an organization. In most of the cases, any organization that is using network depends on the Ethernet technology. In a hub based Ethernet network, when the source wants to send a data packet to destination it broadcasts the message on to the network. Then this packet moves to all the computers connected in the network. Each machine is supposed to ignore the packet if it is not destined for the Internet Protocol (IP) address assigned to that computer/machine. The network interface card (NIC) performs this filtering operation. The packet sniffer is a program that puts the NIC in a special mode called promiscuous mode. In this mode, the NIC does not perform the filtering operation and passes all the received data to the operating system for further processing [3]. The sniffer in the network can be shown in Fig. 1.

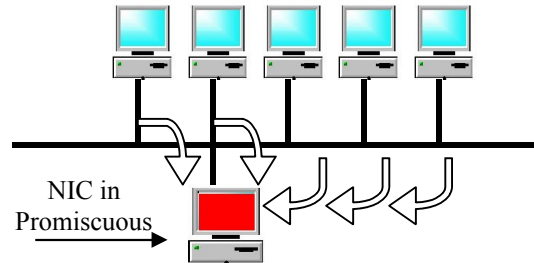


Figure 1. NIC working in Promiscuous Mode

There are many popular sniffers, which are available for free on the internet, as listed below:

- Wireshark
- Kismet
- Tcpdump
- Cain and Abel
- Ettercap
- EtherApe

For sniffing data over secure socket layer, we are considering **Ettercap**. It is a free sniffer tool for UNIX environment but now it is also available for windows based systems.

II. SECURE SOCKET LAYER & SNIFFING

In this section, the method of sniffing over secure socket layer is discussed. Before going into the details of sniffing, working of Secure Socket Layer (SSL) should be discussed. Netscape designed the secure socket layer protocol for web security purpose in 1993.

SSL is a separate protocol layer just for security. It was inserted between HTTP and TCP layer of standard protocol. It can be shown in Fig.2 as:

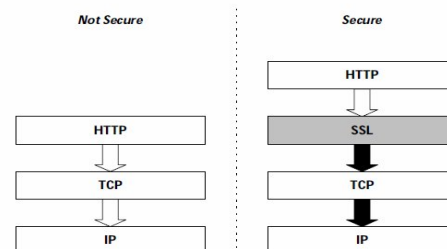


Figure 2. SSL Layer between HTTP and TCP

The SSL protocol consists of a set of messages and rule about when to send (and not to send) each one.

The SSL defines two different roles for the communicating parties. One system is always a client, while the other is a server. The client is the system that initiates the secure communication; the server responds to the client's request. SSL works through a combination of programs and encryption/decryption routines that exist on the web server computer and in web browsers (like Netscape/Firefox and Internet Explorer) used by the Internet public. The process can be shown in Fig.3:



Figure 3. SSL Process

The SSL certificate is installed on a system to encrypt sensitive data such as credit card information. SSL Certificates give a website the ability to communicate securely with its web customers. Without a certificate, any information sent from a user's computer to a website can be intercepted and viewed by hackers and fraudsters. It is similar to the difference between sending a post card and a tamper proof sealed envelope [7].

As discussed earlier, the server installed a certificate in client's system. The **Ettercap** can be used to sniff data over the secure socket layer. Ettercap is a tool made by Alberto Ornaghi (ALoR) and Marco Valleri (NaGA) and is basically a suite for man in the middle attacks on a LAN. For those who do not like the Command Like Interface (CLI), it is provided with an easy graphical interface.

Ettercap is able to perform attacks against the ARP protocol by positioning itself as "man in the middle" and, once positioned as this, it is able to:

- Infect, replace, delete data in a connection
- Discover passwords for protocols such as FTP, HTTP, POP, SSH1, etc ...
- Provide fake SSL certificates in HTTPS sections to the victims.

Man in the Middle Attack:-

This is an attack where a pirate put its machine in the logical way between two machines speaking together as shown in the Fig.4 below.

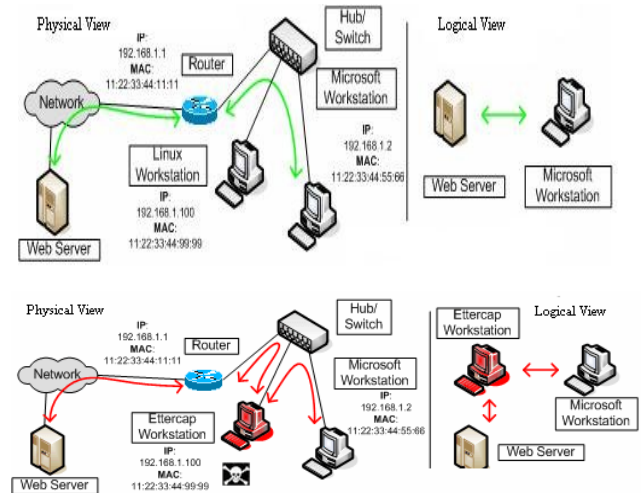


Figure 4. Normal Operation & MITM Attack

Once in this position, the pirate can launch a lot of different very dangerous attacks because he/she is in the way between to two normal machines.

We'll only be able to sniff a network on the same subnet as us. The subnet is usually 255.255.255.0 so click on Options >> Set Netmask and enter the subnet of your network. Now let's start sniffing. Click Sniff >> Unified Sniffing and enter the network interface you want to use. Now we need to scout for hosts on the network. Click on Hosts >> Scan for hosts and wait for it to finish. Then click Hosts >> Host List. This will display a list of hosts. Now you need to define targets for the MITM attack. The router should be added to Target 1 and any other hosts you want to ARP poison should be added to Target 2. This is done by clicking on the host then clicking on either Target 1 or Target 2. Once you've defined your hosts, we need to ARP poison them before we start sniffing [10].

Click on Mitm >> Arp poisoning... to begin.

In the next dialogue be sure to check Sniff Remote Connections (or we won't be able to), then click OK. Now we can start sniffing. Click Start >> Start sniffing to begin.

III. SNIFFING DETECTION

The following methods can be used to detect the sniffer present on the network.

A. Ping Method

In a TCP/IP (IP Version 4) network, every computer has a 32-bit IP address that is used to identify the computer uniquely. Ethernet devices have a 48-bit hardware address, and some kind of mapping between IP and Ethernet is needed when two computers needs to talk to each other. This mapping

is called ARP and is short for Address Resolution Protocol. The 48-bit hardware address is called a MAC-address (Media Access Control) and is often written in hexadecimal format. Using these facts we transmit an "ICMP Echo Request" (ping) with correct IP address and a fake MAC address. Under the normal operation, No one should reply this Request because the MAC address does not match with any computer. But if any computer/NIC working in promiscuous mode will collect this request and reply this request. In this way we can detect that any system is performing sniffing or not. But unfortunately operating system may use virtual MAC address. In this case this technique will not work [4].

B. ARP Method

Network sniffer does not send any packet to the network, so it is hard to detect sniffer. But the behavior of NIC is different from the normal mode. It forwards all the received packets to the operating system or kernel. So in this case hardware filter does not work. We can easily understand the working of this method using a real life example: Imagine a classroom with students and teacher. One student named "Mr. X" came late to class and now he is sniffing the lecture going on in the class room. He listens all the conversations going on in the class room. At the time of attendance if name of sniffer "Mr. X" is called and the "Mr. X" makes a mistake by responding "Present Sir". So NIC in promiscuous mode receives all the packets including those that are not targeting to it, it may reply to a packet which should be filtered by NIC [5] [6]. Now using this technique we can detect a sniffer present on the network. A computer system may set hardware filter in the following mode:

- Unicast
- Broadcast
- Multicast

In ARP, when a nodes wants to know the hardware address of node X, it compose an ARP request packet having (FF-FF-FF-FF-FF-FF) in destination hardware address field [8]. It shows that it is a broadcast message. So all the nodes in the network will receive this packet and only targeted node will reply in normal mode. The encapsulation of ARP message in an Ethernet frame can be represented using this Fig.1-

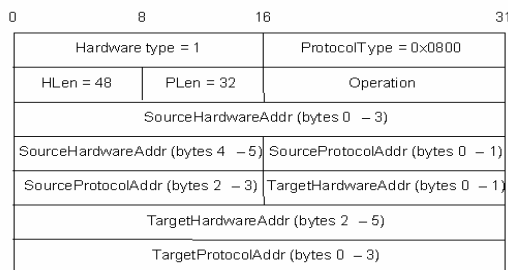


Figure 5. ARP Packet Format

For sniffer detection we set destination or Target hardware address different from the broadcast address. Suppose we set it to 00-00-00-00-00-02. Now in normal mode every node will discard this packet due to hardware filter. In promiscuous mode, the system kernel assumes that it is an ARP request for system so it responds back to the requesting node. In this way we can detect a node for sniffing [2].

C. Decoy Method

As we know many protocols allow plain text passwords and these passwords may be hacked by hacker, who is running the sniffer. The decoy method uses this activity for detecting the sniffer. We set a client and a server using POP, Telnet or any other plain text protocol. We configure some special accounts or virtual accounts on this server. When hacker gets username and password of this account then he tries to log in using this information. We can use standard intrusion detection system to track or log this activity. We can also identify the hacker's system when he tries to log in using that fake username and password. So the decoy method basically works on the principle of Honeypots in which we attract the hacker or intruder, so that we can identify them when they perform any action.

IV. CONCLUSION

In this way it can be concluded that network sniffing is a major threat for computer security because sniffer is a passive component and it does not send any packet to the network. So it is difficult to detect the sniffer. The one solution to this problem is secure socket layer. But data can be hacked over SSL networks using sniffing tools like Etrrcap etc. Similarly sniffer detection methods can be used to detect the sniffers present on the network. All the methods described here may not work with 100% efficiency because the whole paradigm is changing very frequently and the hackers and intruders are discovering new methods for the intrusion. In the similar way new methods should be discovered for security.

REFERENCES

- [1] <http://www.securitysoftwaretech.com/antisniffing>, (2004).
- [2] H. M. Kortebi AbdelallahElhadj, H. M. Khelalfa, An experimental sniffer detector: Snifferwall, (2002).
- [3] Thawatchai Chomsiri, Sniffing packets on lan without arp spoofing, Thrid 2008 International Conference on Convergence and Hybrid Information Technology(2008).
- [4] D. Wu and F. Wong, Remote sni er detection, Computer Science Division, University of California, Berkeley (1998).
- [5] Daiji Sanai, Detection of promiscuous node using arp packets, www.securityfriday.com (2001). 50-51
- [6] DETECTION and PREVENTION OF ACTIVE SNIFFING ON ROUTING PROTOCOL, Pathmenanthan ramakrishna' and mohd aizaini maarof, Student Conference on Research and Development Proceedings, Shah Alam, Malaysia (2002).

- [7] www.evsslcertificate.com/ssl/description-ssl.html
- [8] <http://www.tcpdump.org>.
- [9] <http://reptile.rug.ac.be/~coder/sniffit/sniffit.html>
- [10] www.scribd.com/doc/29844162/Ettercap-Tutorial
- [11] S. Grundschober, Sniffer detector report, IBM Research Division, Zurich Research Laboratory, Global Security Analysis Lab (1998).
- [12] B. Issac S. Kamal, Analysis of network communication attacks, The 5th Student Conference on Research and Development (2007).

AUTHORS PROFILE

Mr. Amit Mishra is working as an Associate Professor in Faculty of Engg. & Tech., Jodhpur National University Jodhpur. His research Intrests include Information Security, Network and Protocols and Data Hacking Analysis.

Secure Communication with Flipping Substitute Permutation Algorithm for Electronic Copy right Management System

¹C. PARTHASARATHY
Assistant professor in IT Dept

²G.RAMESH KUMAR
Assistant Professor

³Dr.S.K.SRIVATSA
Senior professor

¹Sri Chandrashekhendra Saraswathi Viswa Mahavidyalaya University,
Enathur, Kanchipuram – 631 561, sarathy286089@rediffmail.com,

²Department of Computer Science & Applications, Adhiparasakthi College of Arts & Science
G.B.Nagar, Kalavai - 632 506, Vellore District. Tamil Nadu, grk92804@rediffmail.com

³St. Joseph's College of Engg, Jeppiaar Nagar, Chennai-600 064 profsks@rediffmail.com

Abstract-The main objective of this paper is to detect the existence of secret information hidden within an image. Cryptography is one of the most interested and important area in the computer industry that deals with secures transmission of information. Encryption, the process which helps for such secure transmission prevents hackers to access the information. And decryption helps to retrieve the original information. Cryptography provides many methods and techniques for secure communication. Currently there are many industry standard encryption/decryption algorithms including RSA, Rijndael, Blowfish and so forth. However, they are fairly complex and require that one spend a lot of time to comprehend and implement them. This paper introduces simple Encryption/decryption algorithm that is fast and fairly secure. The algorithm manipulates a 128-bit input using flipping, Substitution, and Permutation to achieve its encryption/decryption.

Keywords - Cryptography, Hacker, Security, attack Steganography, Watermarking, compression, authentication.

I. INTRODUCTION

Steganography is a Greek word meaning covered or hidden writing. It is the art and science of secret communication, aiming to conceal the existence of the communication. This is a different from Cryptography, where the existence of the communication is not disguised but the message is obscured by scrambling it. Use of cryptography would not stop a third party knowing that some secret communication is going on. In steganography, the message to be sent is concealed in such a way that an intruder would not know whether any secret communication is going on or not. Hiding information inside digital carriers is becoming popular. A rapid growth in demand and consumption of multimedia has resulted in data hiding techniques for files like audio (.wav), images (.bmp, .pnm, .jpg).

Digital images are most common sources for hiding message. The process of hiding information is called an embedding.

Still and multi-media images are subject to transformations for compression, steganographic embedding and digital watermarking. We propose new measures and techniques for detection and analysis of steganographic embedded content. We show that both statistical and pattern classification techniques using our proposed measures provide reasonable discrimination schemes for detecting embeddings of different levels.

Many algorithms are developed for encryption and decryption which provides high security. All these algorithms are kept open to the public and the secrecy of the algorithm lies entirely in the key. This paper stands different that the development of algorithm addresses the user needs in specific, thereby offering more flexibility. With the help of socket program, establish a connection between client and server. Different segments of secret picture were passed as file objects to the server from client.

II. PROBLEM DEFINITION - PROPOSED ALGORITHM

Secure communication with the help of FSP algorithm as follows:

- Step 1: Set the flipping bit.
- Step 2: Change the characters according to the flipping bit.
- Step 3: Check the ASCII table and find keys.
- Step 4: With the help of the keys, make a square matrix, using inverse table.
- Step 5: Do flipping operation.
- Step 6: Repeat the steps 2 to 5.

In Fig 1, PT is the Plain text and CT is the Cipher text.

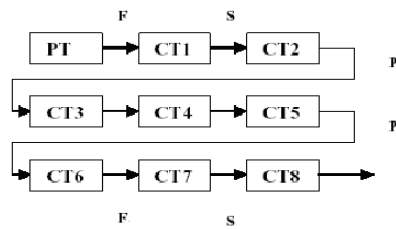


Figure 1. Encryption with 8 levels

A. Flipping Operation

One piece of the secret information is the flipping key and its length is 128 bits, and it is used to obscure the plaintext or cipher text further. Given a 128-bit input PT (Plain Text) and a flipping key F, We denote the flipping operation on PT as below:

$$\text{Output} = \text{Flip (F, PT)}$$

In the flipping operation, its 128-bit input is disguised as follows: For each bit of the input, if the corresponding bit of the flipping key is 0, the corresponding bit of the flipping key is 1, the corresponding output bit will be the complement of the input bit. That is, if the flipping key bit is 0 and the input bit is 0/1, the output of the flipping operation is 0/1. On the other hand, if the flipping key bit is 1 and the input bit is 0/1 the output of the flipping operation is 1/0. In reconstructing the original input, the output of the flipping operation is flipped against the same flipping key.

B. Substitution Operation

This algorithm uses substitution and Inverse Substitution table for encryption and decryption. These tables are generated based upon the ASCII code and the key. Let PT be the plain text, CT be the Cipher text and Key be the Flipping key. In this, plain text as a text file. This file will have all the ASCII characters. The ASCII characters are given in the Table 3. In this, the rows indicate the left digit and the column indicates the right digit. Again this table is subdivided into subsets. For dividing the subset into blocks, we have to follow the following procedure. If the no of characters is less than or equal to 10, we have to divide this into two halves. If the number of characters is even number, we divide it into equal halves. Suppose, the number of characters is odd number, we have to divide this into 2 subsets but the size of the first subset is greater than the second subset by 1. To construct the substitution table 2, it uses key and it will be informed to the receiver in a secure manner.

$$K = \begin{matrix} 15 & 7 & 14 & 6 & 13 & 5 & 12 & 4 & 8 \\ 1 & 9 & 2 & 10 & 3 & 11 & & & \end{matrix}$$

[Numbers 1 to 15 occurring in the key corresponding to the above table]

Using the above key, Flipping key is determined. So the length of the Flipping key is 128bit (ie, $16 \times 8 = 128$).

And Using this key the substitution table and Inverse substitution table is also constructed.

Table 1 – ASCII Table

	0	1	2	3	4	5	6	7	8	9
3			Blank	!	"	#	\$	%	&	'
4	()	*	+	,	-	.	/	0	1
5	2	3	4	5	6	7	8	9	:	;
6	<	=	>	?	@	A	B	C	D	E
7	F	G	H	I	J	K	L	M	N	O
8	P	Q	R	S	T	U	V	W	X	Y
9	Z	[\]	^	_	`	a	b	c
10	d	e	f	g	h	i	j	k	l	m
11	n	o	p	q	r	s	t	u	v	w
12	x	y	z	{		}	~			

Again this table is divided into subsets.

Table 2 – Subset Table

S.No.	Type	Code	No. of Characters
1.	Special Characters	32 – 47	16
2.	Digits	48 – 57	10
3.	Special Characters	58 – 64	7
4.	Uppercase Letters	65 – 90	26
5.	Special Characters	91 – 96	6
6.	Lower case Letters	97–122	26
7.	Special Characters	123–126	4

Table 3 – Block Table

Block No.	Characters	No. of Characters	Block No.	Characters	Block No.
1.	Blank ! @ # \$ % & ,	8	9	U V W X Y Z	6
2.	() = + , - , /	8	10	[] ^ _ `	6
3.	0 1 2 3 4	5	11	a b c d e f g	7
4.	5 6 7 8 9	5	12	h i j k l m	6
5.	: ; < = > ? @	7	13	n o p q r s t	7
6.	A B C D E F G	7	14	u v w x y z	6
7.	H I J K L M	6	15	{ } ~	4
8	N O P Q R S T	7			

Table 4 – Substitution Table

	0	1	2	3	4	5	6	7	8	9
3			55	56	67	58	59	60	61	62
4	69	70	71	72	73	74	75	76	83	84
5	85	86	87	43	44	45	46	47	30	31
6	32	33	34	35	36	16	17	18	19	20
7	21	22	4	5	6	7	8	9	48	49
8	50	51	52	53	54	63	64	65	66	67
9	68	77	78	79	80	81	82	88	89	90
10	91	92	93	94	37	38	39	40	41	42
11	23	24	25	26	27	28	29	10	11	12
12	13	14	15	0	1	2	3			

Table 5 – Inverse Substitution Table

	0	1	2	3	4	5	6	7	8	9
0	123	124	125	126	72	73	74	75	76	77
1	117	118	119	120	121	122	65	66	67	68
2	69	70	71	110	111	112	113	114	115	116
3	58	59	60	61	62	63	64	104	105	106
4	107	108	109	53	54	55	56	57	78	79
5	80	81	82	83	84	32	33	34	35	36
6	37	38	39	85	86	87	88	89	90	40
7	41	42	43	44	45	46	47	91	92	93
8	94	95	96	48	49	50	51	52	97	98
9	99	100	101	102	103					

C. Permutation Operation Proposed Folding Technique

The origin of folding is from paper folding nature. This folding is broadly divided into 3 angles of processing:

1. Vertical Folding
2. Horizontal Folding
3. Diagonal Folding

Consider there are 5 rows present in the plain text document. Cipher text created with respect to

1 ↔ 5
2 ↔ 4
3 ↔ 3

Note : Exchange occurs

the horizontal folding method finds the mid-row of whole text. With respect to that mid row subsequent rows are exchanged.

PLAIN TEXT					CIPHER TEXT				
A	B	C	D	E	E	D	C	B	A
F	G	H	I	J	J	I	H	G	F
K	L	M	N	O	O	N	M	L	K
P	Q	R	S	T	T	S	R	Q	P
U	V	W	X	Y	Y	X	W	V	U

Figure 2.Vertical Folding Technique

PLAIN TEXT					CIPHER TEXT				
A	B	C	D	E	U	V	W	X	Y
F	G	H	I	J	P	Q	R	S	T
K	L	M	N	O	K	L	M	N	O
P	Q	R	S	T	F	G	H	I	J
U	V	W	X	Y	A	B	C	D	E

Figure 3.Horizontal Folding Technique

PLAIN TEXT					CIPHER TEXT				
A	B	C	D	E	A	F	K	P	U
F	G	H	I	J	B	G	L	Q	Y
K	L	M	N	O	C	H	M	R	W
P	Q	R	S	T	D	I	N	S	X
U	V	W	X	Y	E	J	O	T	Y

Figure 4.Diagonal Vertical Folding Technique

In the case of vertical folding method columns are exchanged dynamically. It is same as horizontal folding using column processing instead of row processing.

The diagonal folding method must be implemented in square matrix arguments. If not proper padding must be added to get the appropriate solution. On the side of decryption padding must be eliminated after processing.

```
for(int i=0;i<5;i++)
    for(int j=0;j<5;j++)
    {
        int p=(i*10)+j;
        for(int k=0;k<5;k++)
            for(int q=0;q<5;q++)
            {
                if(p==a[k][q])
                    b[i][j]=((k+1)*10)+q+1;
            }
    }
```

Program 1.Substitution - forming inverse table

D. Encryption Level

The last piece of the secret information is the encryption level. It is a positive integer. The higher the encryption level is, the more secure the algorithm is. However, we should be cautious with large values of the encryption level since the increasing of the encryption level is proportional to the decreasing of the Encryption / decryption speed.

E. Traffic padding

Effective countermeasure to traffic analysis is traffic padding. Traffic padding is one of the functions of link encryption approach. It produces cipher text output continuously in the picture; even in the absence plaintext a continuous random data stream is generated. When plaintext is available, it is encrypted and transmitted. When input plaintext is not present, random data are encrypted and transmitted. It shown in figure 5,

Advantage of traffic padding:

- It is impossible for an attacker to distinguish between true dataflow and padding data flow.
- It is impossible to deduce amount of traffic.

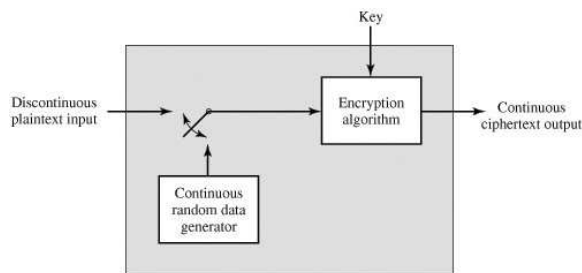


Figure 5. Traffic padding

III. BRUTE FORCE ATTACK

To hack into the FSP encryption/decryption algorithms using the brute force approach, one needs to guess the flipping key, the Substitution function, the permutation function and the encryption level.

A force attack or exhaustive key search is a strategy that can in theory be used against any encrypted data by an attacker who is unable to take advantage of any weakness in an encryption system that would otherwise make them task easier. It involves systematically checking all possible keys until the correct key is found. In the worst case, this would involve traversing the entire search space

A. The Number of the Flipping Keys

The resources required for a brute force attack scale exponentially with increasing key size, not linearly. As a result, doubling the key size for an algorithm does not simply double the required number of operations, but rather squares them.

There are 128 bits in a key. Each bit can be either 1 or 0. Therefore, there are 2^{128} flipping keys.

IV. LAN CONNECTION

The program or process initiating the communication is called a client process, and the program waiting for the communication to be initiated is the server process. The client and server processes together form a distributed system.

Step 1: Start.

Step 2: Select the image file.

Step 3: Encode the information into the image file.

Step 4: Pass the image, to image splitter application, enter the number of segments as input. Multiple image files will be created.

Step 5: Using socket programming, establish a connection between client and server.

Step 7: Different segments were passed as file objects to the server after connecting to the server.

Step 8: Stop.

```
public static void main(String[] args) {
    // TODO Auto-generated method stub
    try
    {
        File file = new File("C:/test.jpg");
        InputStream fis = new
        FileInputStream(file);
        long fileLength = file.length();
        long numberOfSplits = 5;
        long splitFileSize =
        fileLength/numberOfSplits;
        byte[] byteArray =
        new byte[(int)splitFileSize];
        System.out.println("length of the
        file::"+fileLength);
        System.out.println("split file
        size::"+splitFileSize);
        fis.read(byteArray, 0,
        (int)splitFileSize);
        File file2 = new
        File("C:/test1.jpg");
        OutputStream fos = new
        FileOutputStream(file2);
        fos.write(byteArray);
        System.out.println("length of file
        2::"+file2.length());
        fis.close();
        fos.close();
    }
}
```

Program 2. Split the image

V. Description Diagram for Watermarking

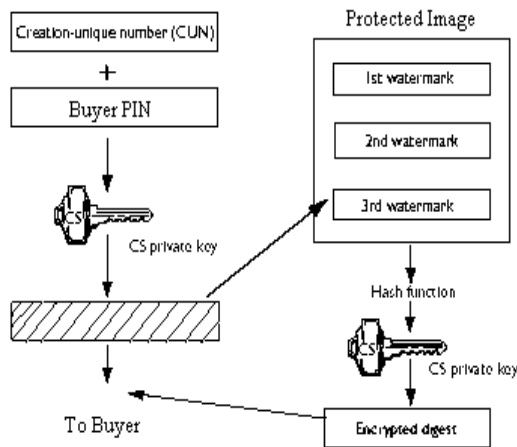


Figure 6. Water marking in Protected Image

The media distributor inserts the third watermark, which contains the document Creation Unique Number (CUN) and the buyer's PIN encrypted with the collecting society's private key.

VI. IMPLEMENTATION DETAILS

This paper consists of implementing the Electronic Copyright Management System (ECMS). In ECMS there are four modules.

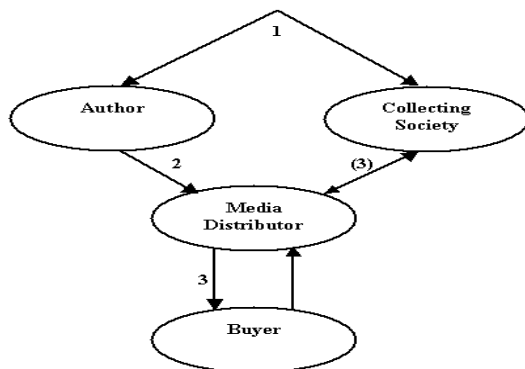


Figure 7. Four modules of ECMS

In Author Module Creation Unique Number is embedded into the Image using author private key. In the embedding of CUN it uses asymmetric watermarking algorithm. Distributor PIN is also embedded into the image using private key Asymmetric encryption algorithm.

Collection Society is the trusted third party that will ensure that the protected document traded correctly. It involves transaction between buyer & media distributor.

In collection Society module, Buyer PIN is embedded into the image using CS private key in Asymmetric encryption. It also computes Hash value of the image which should be sent to buyer. It is used for authentication purpose. This hash value is also appended into the image and the encrypted image is transferred to the buyer using LAN or Email networks.

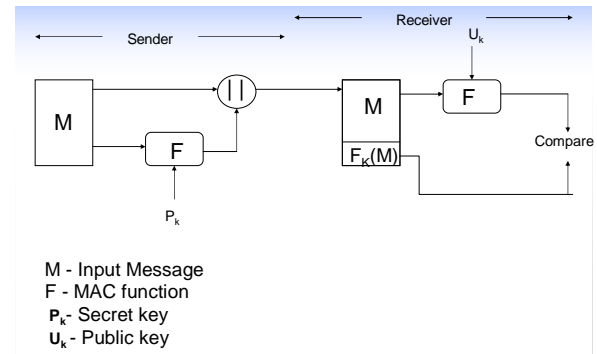


Figure 8. Message authentication

In Buyer module, Buyer decrypts the encrypted digest using CS public key and the digest value is computed. Hash value is recomputed from the decrypted digest and the hash value is compared. If these values are same then it ensures no transmission loss. From third encrypted watermark buyer decrypt the Buyer PIN from it and ensures it legal ownership. Control Authority is used for Illegal usage detection phase. It compares CUN with buyer watermark and distributor watermark and detects the legal or illegal ownership.

A. Algorithm - Server Side:

Sockets are interfaces that can "plug into" each other over a network. Once so "plugged in", the programs so connected communicate. A "client" program can then connect its own socket to the server's socket, at which time the client program's writes to the socket are read as stranded input to the server program, and stranded output from the server program are read from the client's socket reads.

Step 1: Different segments were received as file objects.

Step 2: Using Image Merger application, the segments are merged back to a single file.

Step 3: Apply the FSP algorithm Decode the information.

Step 4: Both the server and client socket connection is closed.

Step 5: Stop.

B. Author Module

In this module CUN and Distributor PIN is encrypted using author private key and the encrypted info is embedded into the image using transaction watermark embedded. In this module all the info embedded into the watermarked image is decrypt and decoded using transaction watermark decoder.

In our approach, the document is self-contained. At any given instant it contains all the information needed to verify whether the current holder is using the data legally. No attempt is made to trace the document history, however, either by watermarking the document each time the owner changes, or by recording transaction details in a register. We take particular care to allow each actor to check that the data exchange was carried out correctly. The basic principle underlying our ECMS strategy is that the data holder's name must be watermarked into the data to prove legal ownership. To ensure that a document is being used legally, any authorized person can check the watermark field the holder's name is written in. We also envision a protocol-level mechanism that addresses the reversibility problem by preventing data holders or counterfeiters from benefiting from watermark removal: at no step of the transaction can a counterfeiter insert a fake watermark, so a counterfeiter cannot prove document ownership. To keep misappropriating persons from writing their names into the data, the ECMS assumes that the seller (or the author when a media distributor sells the document) embeds the watermark.

B. Collection Society Module

In this module Buyer PIN and total document is encrypted using CS private key. If author wants to sell copies of her document through a media distributor, she embeds a second watermark into the document. This watermark contains a personal identification number (PIN) identifying the media distributor, and the document's CUN. Author encrypts the watermark string with her private key and a copy of the encrypted string, which distributor can use to verify that author really inserted his name into the document. Distributor can use Author's public key to read the encrypted string, and watermark detection software to verify it. (Unlike with the first watermark, only an asymmetric cryptography scheme can be used here.)

C. Buyer Module

In this module, buyer verification is achieved by checking the watermarked string with the original watermark using watermark decoder.

String with encrypted third watermark is decrypted using CS public key and the obtained CUN and Buyer PIN is compared

- BUYER passes his PIN to Distributor.
- Distributor passes buyer's PIN, the CUN, and a string with the second watermark's content (that is, Distributor's PIN and the CUN encrypted with author's private key) to the CS.
- The CS passes revenue to Author.
- After encrypting the string with buyer's PIN and the CUN with its private key, the CS embeds the second and the third watermarks into its copy of the document.
- The CS computes a digest of the watermarked document using a proper hash function, signs the digest with its private key, and sends the signed digest and the third, encrypted, watermark to distributor.
- Distributor embeds the third watermark into the document and gives it, the encrypted third watermark, and the signed digest to buyer.

Verification Process: To verify that Distributor has embedded his PIN within the data, Buyer need only decrypt the third watermark using the CS public key. To check whether the CUN embedded in the third watermark corresponds to that in the first, Buyer can compute the digest of the watermarked document and confirm that it corresponds to the digest computed by the CS. Such a digest also allows buyer to verify the integrity of the watermarked document that is he can confirm that Distributor has not modified the original document.

D. Control Authority Module

This phase is used to verify the illegal usage. Protecting Data from Illegal Use Control authority asks buyer to prove his right to a digital document in its possession. Buyer can simply give the watermarked document and the file with the encrypted third watermark to the control authority. The CA first checks the encrypted.

Third watermark for buyer PIN, then, by applying a watermark detection engine to the protected document, it verifies that the watermark with buyer's PIN is actually embedded in the data. Finally, the CA, which knows both the true CUN and author's secret key, can control whether the CUN contained in the third watermark matches the document identity.

Indeed, the CA would not really need the user's file with the encrypted third watermark if it could get this information directly from the CS. Rather than storing all watermarking codes or digests, the CS can simply compute them whenever it needs to, provided the CA gives it the required information. In particular, the CS can generate the second and third watermark and the digest if it knows the media distributor's PIN, the buyer's PIN, the CUN, and the author's identity.

VII. RESULTS AND DISCUSSION

Here the new variant FSP Algorithm developed has been adopted successfully to implement watermarking technique used for invisible information retrieval hidden in a picture message in ECMS. The secret information sending / retrieval among the four modules of ECMS are carried out and the result obtained is satisfactory as shown in the figure 11. The image format BMP is used for embedding the information.

At encryption level the newly developed FSP Algorithm helps to encrypt the incoming information and the spatial domain technique converts it into watermarked image is shown in the figure 9. The visible image now contains the secret information in invisible mode.

At decryption level the watermark decoding process which again uses the FSP Algorithm gets back the secret information in its original form. The decoding with asymmetric watermarking technique is illustrated shown in figure 10. The algorithm is fast as it uses 128 bits length flipping key and works fairly secure, since no unauthorized person can in anyway access the secret information as they require integrating application knowledge which is available to only authorized and intended receivers.

A. Conclusions

This paper addressed the problem of Copyright protection in open network environments. Author Module embeds the CUN and Distributor PIN in the image. In this module FSP algorithm is used to generate public and private keys. CS module embeds the Buyer's PIN into the image using CS private key. Hash value of the image is computed using hashing algorithm. It helps for authentication purpose. In buyer module hash value of the received image is computed using hash function. Buyer confirmation phase is used for authentication purpose. CA module detects illegal usage. Image file is transferred via LAN or email. This proposed scheme may further be

enhanced and to be used in Copyright protection. In addition to that all the image formats should be supported by the software and the e-commerce used in e-transaction will be added in future. This software needs facility of Monitoring and analyzing intruders and raising alarm with new technique. The FSP encryption / decryption algorithm is a simple algorithm based on the flipping, substitution and permutation operations. It is fast and fairly secure. However, it is only suitable for applications that do not expose the inputs and the encrypted form of the inputs to the public. If there is a need for the applications to expose its inputs and its encrypted forms of the inputs, then it should use the FSP encryption / decryption algorithm instead.

REFERENCES

- [1] Whitfield diffie and Artine E.Hellman "New directions in Cryptography",IEEE Transactions on Information Theory, Vol IT-22, no 6, November 2010.
- [2] J. Fridrich, M. Goljan and R. Du, "Reliable Detection of LSB Steganography in Color and Gray-Scale Images," in Magazine of IEEE Multimedia Special Issue on Security, October-November 2001, pp. 22-28.
- [3] N. Jacobsen, K. Solanki, U. Madhow, B. S. Manjunath and S. Chandrasekaran, "Image-adaptive high-volume data hiding based on scalar quantization," in Proc. IEEE Military Communications Conference (MILCOM), Anaheim, CA, USA, Vol. 1, pp 411-415, Oct. 2002.
- [4] A. Westfeld, "High Capacity Despite Better Steganalysis (F5A Steganographic Algorithm)", in LNCS Vol.2137, Springer-Verlag, New York Heidelberg Berlin, pp. 289302, 2001.
- [5] Alessandro Piva and Franco Bartolini, University of Florence, Mauro Barni University of Siena, "Copyright protection in Open networks", IEEE Internet Computing, June2002.
- [6] Kalaichelvi. V & RM.Chandrasekaran,18-20 December. 2008, "FSP Algorithm for encryption/decryption", ICCCN 2008, Proceedings of the International conference on computing and communication network, Karur, Tamilnadu, pp .245-251.

- [7] Rade Petrovic, 2003, "Copyright Protection based on Transaction watermark" IEEE, October 1-3, 2003, Vol 2, 0-7803-7963-2, pp.509-518, Telecommunications in Modern Satellite, Cable and Broadcasting Service, San Diego,CA,USA.
- [8] Richard Smith, 2002, "Internet Cryptography", Pearson Edn. Pvt. Ltd, 2nd Edition, ISBN 0-201-92480-3, pp. 39- 44, Boston, MA 02131 USA.
- [9] T. Furon, I. Venturini, and P. Duhamel, 2001, "Unified Approach of Asymmetric Watermarking Schemes",2002, Security and Watermarking of Multimedia Contents III, P.W. Wong and E. Delp, eds., Proc. SPIE, vol. 4314, 2001, pp. 269-279.
- [10]Tsuhan Chen, Kou-Sou Kan and Ho-Hsun Chang, 2003, "Watermarking 2D/3D Graphics for Copyright Protection", IEEE ICASSP 2003. Vol: 4, pp: IV- 720-3, ISBN: 0-7803-7663-3, Internet & Multimedia Lab., Chunghwa Telecom Labs, Taoyuan, Taiwan, 7810434, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1202744.
- [11]Wei Li, Xiangyung Xue, and Peizhong Lu, Dept of CSE, University of Fudan, Shanghai, China, 2003, "A Novel Feature-based Robust Audio Watermarking for Copyright Protection", IEEE Computers and Communication 2003, ISBN: 0-7695-1916-4,554-560,Washington,DC,USA.
- [12]William Stallings, 2008,"Cryptography and Network Security", Pearson Edn. Pvt. Ltd, 2008, 4th edition, ISBN 13:9780132023221,pp. 26-29, Akhil books Pvt Ltd,India.
- [13]Anderson and Petitcolas 2001 Anderson.R, Petitcolas.F, "On the limits of the steganography", IEEE Journal Selected Areas in Communications, 16, 4,474{481.
- [14]Bassia et al. 2001 Bassia, P., Pitas, I., Nikolaidis, N.: "Robust audio watermarking in the time domain", IEEE Transactions on Multimedia, 3, 2, 232{241.
- [15]Cedric et al. 2000 Cedric, T., Adi, R.,McIoughlin, I.: "Data concealment in audio using a nonlinear frequency distribution of PRBS coded data and frequency-domain LSB insertion", Proc. IEEE Region 10 International Conference on Electrical and Electronic Technology, Kuala Lumpur, Malaysia, 275-278.
- [16]Fridrich et al. 2002 Fridrich, J., Goljan, M., Du, R.: (2002), "Lossless Data Embedding New Paradigm in Digital Watermarking", Applied Signal Processing, 2002, 2, 185{196.
- [17]Yeh and Kuo 1999 Yeh, C., Kuo, C.: , "Digital Watermarking through Quasi m-Arrays", Proc. IEEE Workshop on Signal Processing Systems, Taipei, Taiwan, pp. 456-461. [Zwicker 1982] Zwicker, E.: Psychoacoustics, Springer Verlag, Berlin, Germany.
- [18]A. M. Alattar, G. I. Al-Regib, and S. A. Al-Semari, "Improved selective encryption techniques for secure transmission of MPEG video bit-streams", Proceedings of the IEEE International Conference on image Processing(ICIP'99) vol 4, pp. 256-260, Kobe, Japan, October 1999. IEEE Signal Processing Society.

AUTHORS PROFILE

¹**C.PARTHASARATHY** has been working as a Assistant professor in the Department of Information Technology in Sri Chandrashekhendra Saraswathi Viswa Maha Vidyalaya University, Enathur, Kanchipuram -631 561 since 2006. He has completed his M.C.A from in Madras University, and M.Tech in Sathyabama University and M.Phil in Computer Science from Annaamalai University. Since January 1st 2001 C.Parthasarathy has been a Lecturer in various colleges. He has been research in Network Security. He has been a Ph.D student in network security at the S.C.S.V.M.V University of Kanchipuram. He is currently focusing on the creating a new algorithm in Steganography. He has attended international and National seminars, Conferences, Workshops and also presented papers.

²**G.RAMESH KUMAR** started his career as Lecturer in 1994 and having more than 16 years of teaching experience. He completed his M.Phil degree from Manonmaniam Sundararanar University, Tirunelveli in 2003. He served as a member of the inspection committee for Computer Science courses at Thiruvalluvar University, Vellore. He has been appointed as member in Board of studies in Computer Science and Computer Application(PG) for a period of three years from 16.04.2010 to 15.04.2013. He has written three Text books for Computer Science UG & PG Courses. Presently he is working as Assistant Professor of Computer Science, Adhiparasakthi College of Arts & Science, Kalavai - 632 506. Vellore District. Tamil Nadu.

³**Dr.S.K.SRIVATSA** was born at Bangalore on 21st July 1945. He received his Bachelor of Electronics and Communication Engineering Degree (Honors) from Javadpur University (Securing First Rank and Two Medals). Master Degree in Electrical Communication Engineering (With Distinction) from Indian Institute of Science and Ph.D also from Indian Institute of Science, Bangalore. In July 2005, he retired as professor of Electronics Engineering from Anna University. He has taught twenty-two different courses at the U.G. Level and 40 different courses at P.G. Level during the last 32 years. He has functioned as a Member of the Board of Studies in some Educational Institutions. His name is included in the Computer Society of India database of Resource Professionals. He has received about a dozen awards. He has produced 23 PhDs. He is the author of well over 350 publications.

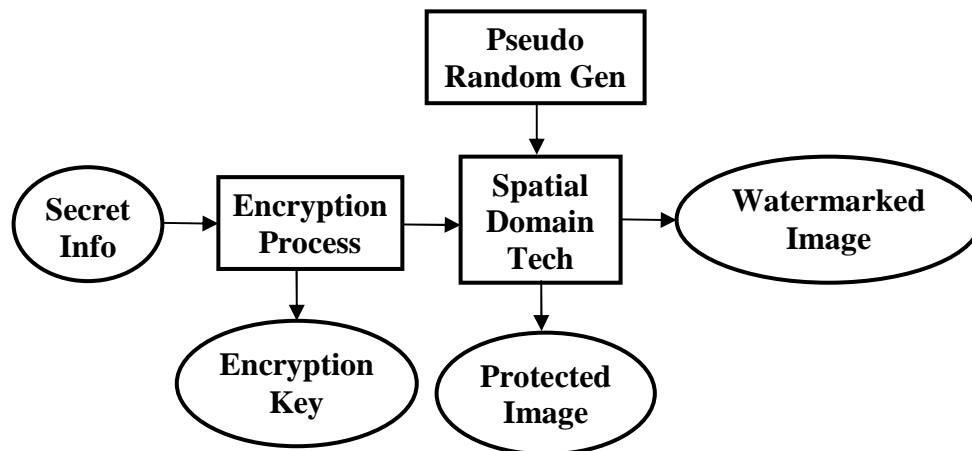


Figure 9. Encoding with Watermarking technique.

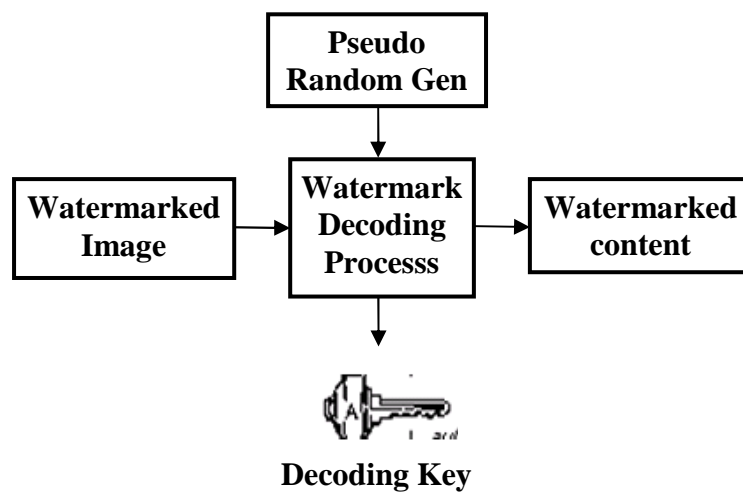


Figure 10. Decoding with asymmetric watermarking technique

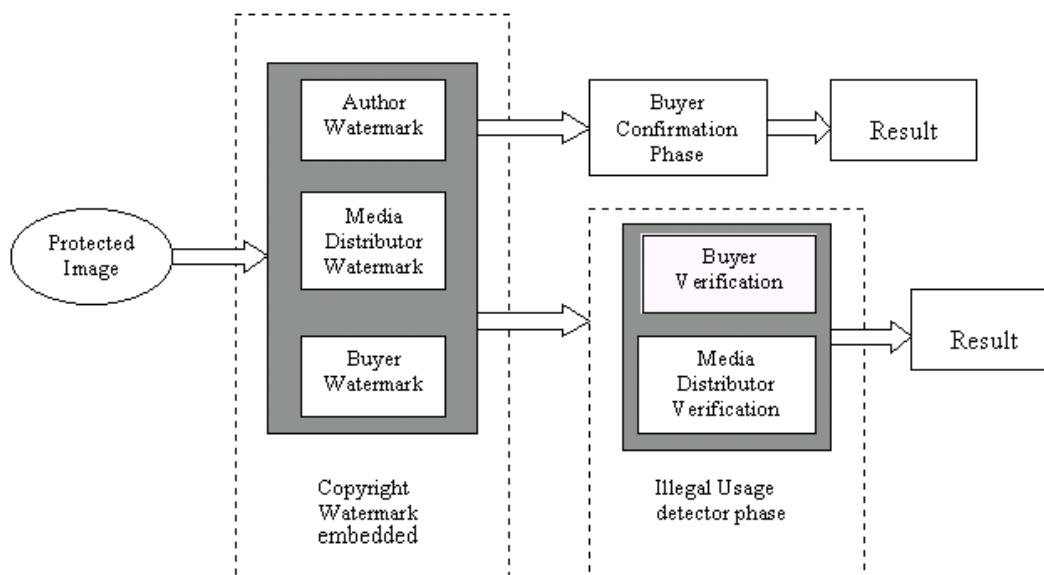


Figure 11. Copyright watermark embedding

Context Based Word Sense Extraction in Text: Design Approach

¹Ranjeetsingh S.Suryawanshi
M.Tech.(Computer Engineering)
ranjeetsuryawanshi06@gmail.com

²Prof.D. M. Thakore
Department of Computer Engineering
dmthakore@bvucop.edu.in

³ Kaustubh S. Raval
M.Tech.(Computer Engineering)
raval_kaustubh@yahoo.co.in

^{1,2,3} Bharati Vidyapeeth Deemed University, College of Engineering, Dhankawadi, Pune – 411043.

Abstract - Today user performs most of his work with electronic document. Due to huge volumes of unstructured electronic texts available, it requires automated techniques to analyze and extract knowledge from these repositories of information. This unstructured text can be available in the form of emails, normal text document and HTML files etc. Understanding meanings and semantics of these texts is a complicated problem. This problem becomes more vital, when meanings with respect to context, have to be extracted.

Text mining, also known as Intelligent Text Analysis, extract interesting information and knowledge from unstructured text. The agent for Context Based Sense Extraction in Text formulates the standard Natural language processing rules with certain weights. These weights for each rule ultimately support in deciding the particular meaning of a word and sentence. In this paper architecture and design of Context Based Word Sense Extraction have been presented.

Keywords- Text Mining, Word Sense, Data Mining.

I. INTRODUCTION

Sensing multiple meanings in a large electronic text is very difficult by a machine as compared to natural human language. In natural language, human extract the word sense by relating it to that particular context. But for electronic text this work is done by natural language processing by extracting two properties of word.

1. It removes ambiguity of an individual word that can be used (in different contexts) to express multiple meanings.
2. It identifies different meanings of word by extracting relation between two words that are spelled the same way.

To sense any word, two resources are necessary: A context in which the word has been

used and Knowledge for finding relation of word in context. Human has an ability to find relation knowledge of word in a context. For example for a word “Fine” in the context of human condition it associate more word like “look”, “well”, “feel”. Due to these associated word it will definitely describe human condition and not refer to penalty.

Due to lack of knowledge intelligence in computer, it uses extra resources to sense word like dictionaries, tagged documents etc. Following are main approaches used in computer to sense word [5].

Dictionary-based Algorithms:

It uses knowledge resources in the form of machine readable dictionaries to extract multiple sense of word. Dictionary defines a term in a particular subject.

Supervised Disambiguation Algorithms:

It uses knowledge resource in the form of tagged corpora which defines meaning of word. It builds classifier which classifies new word correctly depending on their context of use. It needs large sense training set to extract sense of word.

Unsupervised Disambiguation Algorithm:

Unsupervised disambiguation algorithm is equivalent to clustering in which they group instances of a word by meaning.

II. TECHNOLOGY FOUNDATIONS

A. Data Mining:

Data mining is the analysis of large quantities of data, so as to retrieve useful and meaningful patterns and rules. The volume of data is increasing day by day. In order to overcome the deficiencies of manual analysis, data mining techniques can be used, so that an accurate and optimal result is obtained. Data mining involves a series of steps. In classification, the incoming data is grouped by comparing their features to the predefined elements of a class. In estimation, a border limit is established and checked whether the data value is above or below that limit and the classification is done. Association rules helps to decide which combinations are best, so that the outcome is best. In clustering the grouping of data is done. [4]

There is a wide array of techniques that can be used to mine data. Statistical techniques, neural networks, machine learning techniques, genetic algorithms, rough sets techniques, fuzzy set techniques, decision tree building procedures, k nearest neighbor's techniques, and other tools are available for data mining. Each of these techniques has its strengths and weaknesses, and part of the value provided to the project by the data mining team lies in understanding which techniques to use, and when to use.

B. Text mining:

Text mining, also known as Knowledge-Discovery in Text (KDT), refers to the process of extracting interesting information and knowledge from unstructured text. Data mining tools are designed to handle structured data from databases, while Text mining can handle unstructured or semi-structured data sets such as emails, full-text documents and HTML files etc. [1]

Human can easily handle contextual meaning but computer cannot handle easily

spelling variations and contextual meaning of text until some rules provided to the computer. This scenario becomes more significant and critical when the meanings of a piece of text have to be extracted in a particular context. Natural language processing (NLP) is used to determine which sense of a word should be adopted for each instance of a word. Figure 1 depicts a generic process model for a text mining application [1].

Starting with a collection of documents, a text mining tool would retrieve a particular document and pre-process it by checking format and character sets. Then it would go through a text analysis phase, sometimes repeating techniques until information is extracted. The resulting information can be placed as a pattern discovery which will help to interpret target knowledge.

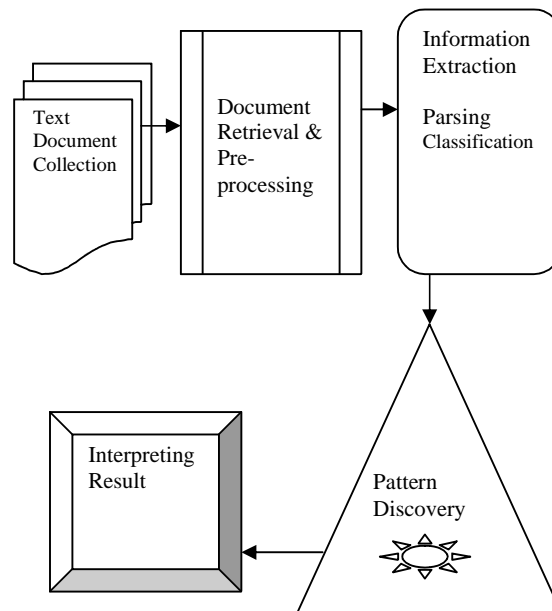


Figure 1: Generic process for a text mining

C. Natural language processing:

Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. Natural language is used to represent human thoughts and human actions. Natural language processing produced technologies that teach computers to analyze, understand, and even generate text. Some of the technologies that have been developed and can be used in the text mining process are information extraction, categorization, clustering, concept linkage, information visualization, and question answering. Applications of NLP include a number of fields of studies, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence etc. [6]

III. SYSTEM DESIGN

Figure 2 shows Use Case Diagram for Context based Word Sense in text with following entity.

Actor:

- 1) User.

Use Cases:

- 1) Frequency Count for Word.
- 2) Calculate Weight Matrix.
- 3) Find out relationship between words.

As shown in figure 2 user will enter query to the system. Then system will generate frequency count for scenario provided by user. If multiple meanings possible for entered query, then system will find relation of word within documents. Finally system calculate weight matrix to rank possible senses of word.

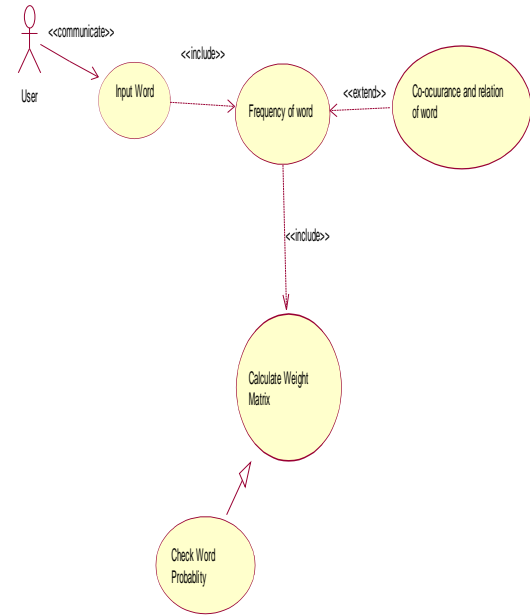


Figure 2: User Interaction

Figure 3 shows the data flow for the 'Context based word sense text-mining system'. The basic process is broken down into sub-processes such as 'Process 1: Parsing, ordering and finding key-phrase', 'Process 2: Deciding contexts', 'Process 3: Calculate Frequency Count for word', 'Process 4: Calculate weight matrix value and associated word senses'.

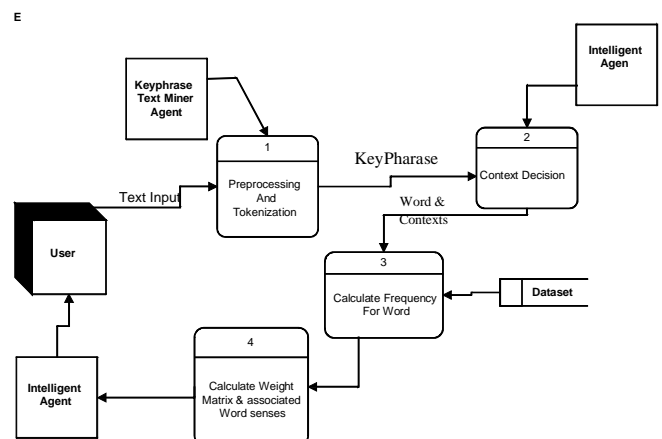


Figure 3: Data flow for Context based word sense system.

IV. SYSTEM ARCHITECTURE

Figure 4 Shows Architecture of Context based Word Sense, which work in three phases.

- A) Find Frequency Count For Word.
- B) Calculate Weight Matrix.
- C) Sentence selection and Contextual meaning of word.

A. Text Structure Analysis and Word Segmentation:

A very first step in a system is to count occurrence of word in a dataset to generate weighted matrix. System will take input as word to be sensed and start searching word frequency in particular contexts. This frequency count is used calculate probability of word sense in given contexts. Then it will group words that have the same conceptual meaning like employee, employer etc. System will perform grouping of words using clustering by calculating word relativity.

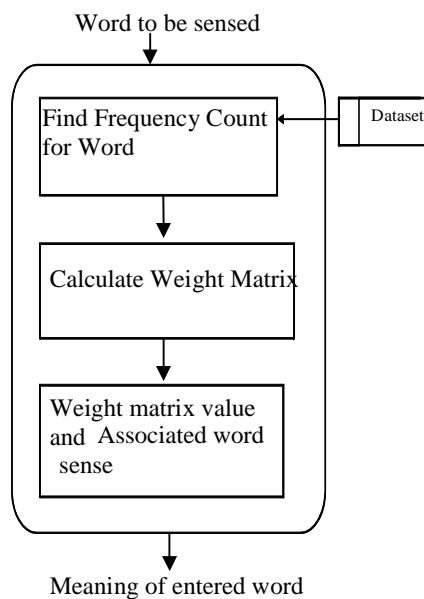


Figure 4: Architecture of Context based Word Sense

Word count probability describes semantic relation of two or more words in a given context.

B. Calculate Weight Matrix :

Depending upon frequency count of word system will generate weight matrix for extracting multiple sense of word. Multiple sense of word depends upon set of co-occurrence between each term and the frequent terms, i.e., occurrences in the same sentences, is counted. For example to extract word sense of “Fine” weight matrix can be generated as follows by 2 X 2 matrix.

	<i>Fine</i>	<i>Fine</i>	<i>Total</i>
Condition	R11=70	R12=115	R1=185
Penalty	R21=50	R22=40	R2=90

Table 1: 2 X 2 Weight Matrix.

As shown in table1 first row describes relation of word “Fine” as satisfactory condition 185 times and second row describes penalty or punishment relation of word “Fine” 90 times.

C. Sentence selection and Contextual meaning of word:

Depending on user request, system will collect list of sentences for given word and will extract best Contextual meanings of word from dataset. To extract best possible sense of word it uses weight matrix and decides rank of sentences.

CONCLUSION

Hence, we can conclude that using word frequency count and weight matrix calculation we can weigh the documents and the system need to incorporate ‘Document Weighing Algorithm’ which will perform this functionality.

REFERENCES:

- [1] Vishal Gupta, Gurpreet Lehal, "A Survey of Text Mining Techniques and Applications," Emerging technologies in web intelligence, vol. 1, august 2009 pages 60-75.
- [2] Hisham Al-Mubaid, "Context-based Term Disambiguation in Biomedical Literature", IEEE Evolutionary Computation, 2006 Pages: 1577 - 1584.
- [3] Yutaka Matsuo, Mitsuru Ishizu, "Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information "National Institute of Advanced Industrial Science and Technology, Japan
- [4] Anupama Surendran, "Data mining Techniques to Analyze the Risks in Stocks/Options Investment", International Conference on Intelligent Agent & Multi-Agent Systems, IAMA 2009 Pages: 1 - 3.
- [5] Roberto Navigli, "Word Sense Disambiguation: A Survey", ACM Computing Surveys, Vol. 41, No. 2, Article 10, Publication date: February 2009.
- [6] Gobinda G. Chowdhury, "Natural Language Processing", Dept. of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK.
- [7] Andreas Hotho, Andreas Nurnberger, and Gerhard Paaß, "A brief Survey of text mining", May 2005.
- 2) Professor D.M.Thakore graduated (B.E – Computer Engineering) from Shivaji University, Sangli, and State – Maharashtra in 1990.
He had pursued his M.E. (Computer) from Bharati Vidyapeeth University College of Engineering, Pune in 2004.
He is currently pursuing his Ph.D. with specialization in subject 'Data Mining/Text Mining' from Bharati Vidyapeeth Deemed University College of Engineering, Pune.
- 3) Kaustubh S. Raval graduated (B.E - Computer Engineering) from Gujarat University, Ahmedabad, and State-Gujarat in the year 2009. Currently pursuing M.Tech. (Computer) with specialization in subject 'Data Mining' from Bharati Vidyapeeth Deemed University College of Engineering, Pune.

AUTHORS PROFILE

- 1) Ranjeetsingh S. Suryawanshi graduated (B.E – Computer Engineering) from Pune University, and State – Maharashtra in the year 2005. Currently pursuing M.Tech. (Computer) with specialization in subject 'Data Mining' from Bharati Vidyapeeth Deemed University College of Engineering, Pune.

An Overview on Radio Access Technology (RAT) Selection Algorithms for Heterogeneous Wireless Networks

J.Preethi

Assistant Professor ,
Department of Computer Science and Engineering
Anna University of Technology, Coimbatore
India

Dr.S.Palaniswami

Professor,
Department of Electrical and Electronics Engineering
Government College of Technology
India

Abstract—Next generation wireless networks (NGWN) will be heterogeneous in nature where different radio access technology coexist in the same coverage area. The coexistence of different RATs require a need for Joint Radio Resource Management (JRRM) to support the provision of quality of service and efficient utilization of radio resources. The Joint Radio Resource Management (JRRM) manages dynamically the allocation and deallocation of radio resources between different Radio Access Technology (RAT). The homogenous Call Admission Control (CAC) algorithms do not provide a single solution to address the heterogeneous architectures which characterize next generation wireless networks. This limitation of homogeneous CAC algorithms necessitates the development of RAT selection algorithms for heterogeneous wireless network. The goal is to select the most suitable RAT for each user. This paper investigates ten different approaches for selecting the most appropriate Radio Access Technology (RAT) for incoming calls among the Heterogeneous Wireless Networks. The advantages and disadvantages of each approach are discussed. This RAT selection works in two steps; the first step is to select a suitable combination of cells among the different RATs. The second step chooses the most appropriate RAT to which the users can be attached and to choose the suitable bandwidth to allocate for the users.

Keywords- Radio Access Technology (RAT) selection, Joint Radio Resource Management (JRRM), Heterogeneous Wireless Networks.

I. INTRODUCTION

Over the past twenty years, a number of different wireless technologies have been developed. The first generation mobile communications systems (e.g. Nordic Mobile Telephony (NMT) and Advanced Mobile Phone System (AMPS)) are based on analog transmission techniques. Analog signals are radio transmissions sent in a wave-like form. A mobile device sends the waves to a base station where they are processed to determine the signals next destination (i.e. another base station, mobile phone, land line phone etc.,) Once the destination is determined, the signal is reconstructed as

accurately as possible into its original wave form by the base station.

The second generation was implemented to improve transmission quantity system capacity and network coverage. In second generation systems (e.g., Personal communication systems (PCS), Global System for Mobile Communication (GSM), Code Division Multiple Access (CDMA), Time Division Multiple Access (TDMA) and Orthogonal Frequency Division Multiplexing (OFDM)) are based on Digital Transmission. 2G is used to transmit voice and it introduced a low volume digital data for mobiles such as Short Message Service (SMS) or Multimedia Message Service (MMS).

In digital systems, more efficient use of the available spectrum is achieved by digital encoding of the speech data. Due to the transition from 2G to 3G, a number of standards have been developed, which are categorized as 2.5G. These are add-ons to the 2G standards and mainly focus on deployment of efficient IP connectivity within the mobile networks.

2.5G is a stepping stone between 2G and 3G cellular wireless technologies, invented for marketing purposes only. 2.5G implements a packet switched domain which includes GPRS (General Packet Radio Service), EDGE (Enhanced Data rates for GSM).

The objective of the third generation (3G) is to provide fairly high speed wireless communications to support multimedia, data and video in addition to voice. 3G includes Universal Mobile Telecommunications systems (UMTS) [1], CDMA2000 based on W-CDMA technologies which provides services like wireless access to the Internet and high data rate applications like real time video transmission. To cope with these, high bandwidth services and the enormous increase in the number of users, a more efficient use of radio spectrum is required [2].

In turn, the perspective of beyond 3G systems is that of heterogeneous networks, which provides wireless services independently of its location in a completely transparent way [3]. The user terminal should be able to pick the best access technology such as Wireless Local Area Network (WLAN), the Universal Mobile Telecommunication Systems (UMTS) and the Global System for Mobile Telecommunication (GSM)/Enhanced Data rate for GSM Evolution (EDGE) Radio Access Network (GERAN) at its current location and use the

access technology seamlessly for the provision of desired service. This leads to the introduction of new Radio Resource Management (RRM) techniques referred to as (JRRM) Joint Radio Resource Management algorithms which manages dynamically the allocation and deallocation of radio resources between different Radio Access Technology (RAT). That is, instead of performing the management of radio resources independently for each RAT, some form of overall and global management of the pool of radio resources can be envisaged.

The coexistence of different RATs require a need for Joint Radio Resource Management (JRRM) to support the provision of quality of service and efficient utilization of radio resources. In heterogeneous wireless networks, different RATs coexist in the same coverage area. The goal is to select the most suitable RAT for each user. In this paper, a comprehensive survey of different RAT selection algorithms for a heterogeneous wireless network is proposed. Section II explains about the architecture of heterogeneous wireless network. Section III presents the benefits of Joint Radio Resource Management algorithms and Section IV proposes RAT selection approach for selecting the appropriate RAT for each user. The section V presents the comprehensive survey of RAT selection algorithms and lastly the conclusions are presented in Section VI.

II. HETEROGENEOUS WIRELESS NETWORK BEYOND 3G

Next generation networks will be heterogeneous where different radio access networks such as WLAN, UMTS, WiMax and satellite networks which is illustrated in the Figure 1. In order to provide the mobile users with the requested multimedia services and corresponding quality of service (QoS) requirements[4], these radio access technologies will be integrated to form a heterogeneous wireless access network. Such a network will consist of a number of wireless network and will form the fourth generation (4G) or next generation of wireless networks. However, each access network provides different levels of QoS, in terms of bandwidth, mobility, coverage area and cost to the mobile users.

III. BENEFITS OF HETEROGENEOUS JOINT RADIO RESOURCE MANAGEMENT ALGORITHMS

Each Radio Access Network (RAN) differs from the others by the air interface technology, cell size, services supported, bit rate capabilities, coverage, mobility support etc., therefore the heterogeneous characteristics offered by the network is considered. As a result, RAT provide further flexibility in the way how radio resources can be managed. This lead to the introduction of RRM. The basic function of Call Admission Control (CAC) algorithm is to decide whether a new handoff call can be accepted into a RAT without violating service commitments [5]. CAC has been used in wired networks and in homogenous wireless networks such as GSM, UMTS, WLAN, Satellite network etc., However, homogenous CAC algorithms do not provide a single solution to address the heterogeneous architecture. This limitation of homogenous CAC algorithm necessitates the development of RAT selection algorithm for heterogeneous wireless network.

Joint Call Admission Control algorithm is one of the JRRM algorithms. Within the JRRM, the initial RAT selection, i.e the allocation of connections to specific RANs at session initiation and the Vertical Handover (VHO) i.e the capability to switch ongoing connections from one RAN to another. These are the key enablers to properly manage the heterogeneous radio access network and become the key for the JRRM functions. The benefits of Heterogeneous Joint Radio Resource Management Algorithms are

- Efficient utilization of radio resources,
- Consistent provisioning of QoS across different RATs,
- Overall stability of network,
- Increase in Operator's revenue and
- Enhancement of users satisfaction.

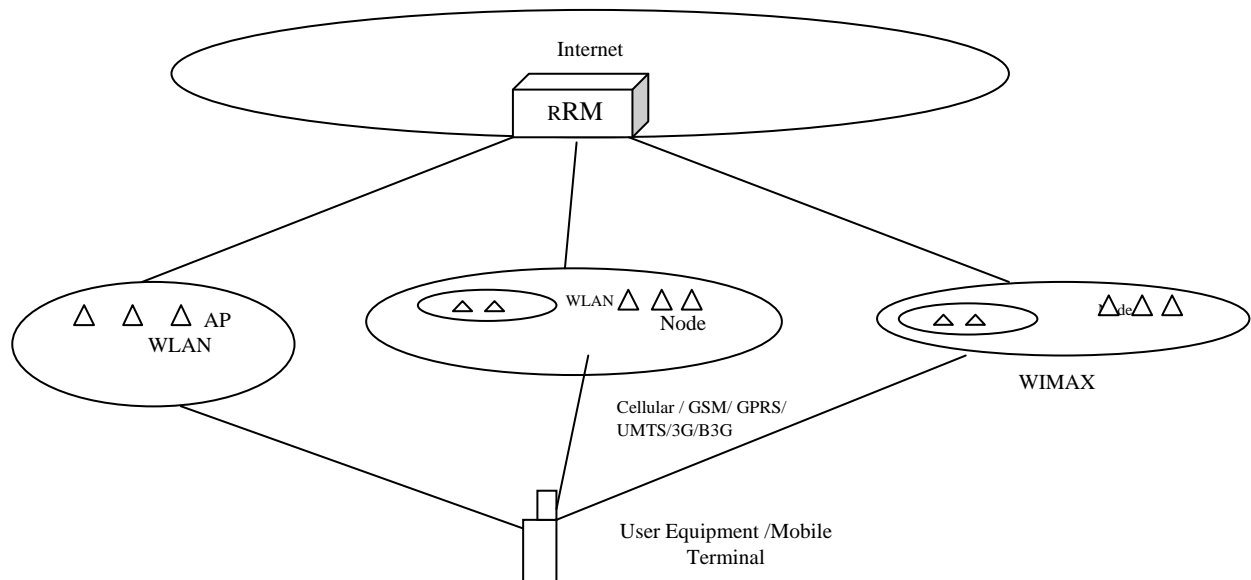


Fig.1. Integration of Heterogeneous Wireless Access Network

IV. APPROACHES TO RAT SELECTION

In this framework [6] as illustrated in Fig. 2, selecting the proper RAT and cell is a complex problem due to the number of variables involved in the decision making process. Furthermore, some of the variables may vary dynamically which makes the process even more difficult to handle.

The RAT selection approaches consists of 3 main parts, Input , RAT selection algorithm and Output . For Inputs, many criteria are considered such as Local RRM, Operator Preferences, User's Preferences, Load conditions, Service type, Service Cost, Interference Conditions. In Decision making block, different RAT selection algorithms are available which are explained in the Section V. The output will give the cell (RAT) to be selected and amount of bandwidth allocated to each RAT). Then the user will be allocated to the selected RAT with the allocated bandwidth.

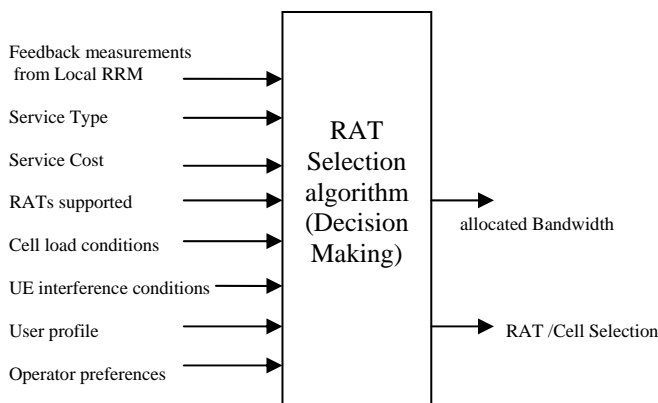


Fig.2. Factors influencing RAT Selection

V. COMPREHENSIVE SURVEY OF RAT SELECTION ALGORITHMS

The section describes ten different RAT selection algorithms for initial RAT selection and Vertical Handover are proposed. References [6],[7] presents the good review of these RAT selection algorithms.

A) Random based RAT selection algorithm

When a new incoming call or vertical handoff arrives, one of the available RAT is randomly selected for the call. If there is no radio resource to accommodate the call in the selected RAT, another RAT is randomly selected. If none of the RAT serves the call, then the incoming call will be blocked/dropped. The advantage of this algorithm is that they are easy to implement. However, they have a high call blocking probability and low radio resource utilization efficiency.

B) Load based RAT selection algorithm

The objective is to uniformly distribute the load among all the available RATs in heterogeneous wireless network. Balancing the load increases the utilization of the radio

resources [8][9]. Load balancing can be forced or unforced. Forced load balancing [10] is carried out by moving some ongoing calls from highly loaded RAT into less loaded RAT, whereas unforced load balancing is achieved only during the new call arrival or in the vertical handoff call.

The major advantage of the load balancing RAT selection network is high network stability, however forced load balancing results in high frequency of vertical handoff calls and high signaling overhead. Load balancing RAT selection algorithms are network-centric and may result in low users satisfaction.

C) Policy based RAT selection algorithm

Policy based RAT selection allocates users to the RAT based on some specific rules specified by the network. A simple policy has been proposed in [11], which includes Voice GERAN (VG) policy, Voice UTRAN (UV) policy and Indoor (IN) policy.

In VG policy, service type is taken as input and allocates voice users into GERAN and other services into UTRAN.

$$f(\text{service}) = \begin{cases} \text{GERAN, if service = voice} \\ \text{UTRAN, if service = www} \end{cases} \quad (1)$$

In VU policy, it acts in opposite direction as VG and allocates voice users to UTRAN and interactive users to GERAN.

$$f(\text{service}) = \begin{cases} \text{UTRAN, if service = voice} \\ \text{GERAN, if service = www} \end{cases} \quad (2)$$

In Indoor(IN) policy, selection is based on whether the user is indoor or outdoor,

$$f(\text{indoor.user}) = \begin{cases} \text{GERAN, if indoor_user=true} \\ \text{UTRAN, if indoor_user=false} \end{cases} \quad (3)$$

Complex policy is proposed in [11] which includes policy like VG*IN, VG*VU, IN*VG policies.

In VG*IN policy, it allocates indoor voice users to GERAN and outdoor data users to UTRAN. Outdoor voice users will be allocated firstly to GERAN to fill the available capacity and then it will direct them to UTRAN. In contrast, the indoor data users will be allocated firstly to UTRAN to fill the available capacity and then it will direct them to GERAN.

In VG*VU policy, it always allocates voice users to GERAN firstly to fill the available capacity and then it will direct them to UTRAN. The data users will be allocated firstly to UTRAN to fill the available capacity and then it will direct them to GERAN.

In IN*VG policy, it always allocates indoor users to GERAN and outdoor data users to UTRAN. Outdoor voice users will be allocated firstly to UTRAN to fill the available capacity and then it will direct them to GERAN. Therefore, indoor data users will be allocated firstly to GERAN to fill the available capacity and then it will direct them to UTRAN.

D) Service-class based RAT selection algorithm

Service-class based RAT admits calls into a particular RAT based on class of service such as voice, video streaming, real-time video, web browsing etc., [12]. This algorithm admits the incoming call that can best suitable for service class of the call. Service-Class based RAT selection has an advantage of high packet-level QoS but they may lead to highly unbalanced network load.

Service-class based RAT can be classified as rigid or flexible. Rigid service class based RAT selection admits an incoming call of specific class into a particular RAT. If the chosen RAT does not provide the enough radio resources for the new call and also if other RATs are not acceptable then the new call will be blocked/dropped. Flexible service class RAT selection attempts to admit an incoming call of a specific class into a particular RAT. If the preferred RAT for this call cannot accommodate the call, other RATs are acceptable. A flexible service-class based RAT has lower call blocking probability when compared to rigid service-class based RAT selection.

E) Service-cost based RAT selection algorithm

Service cost based RAT selection admits incoming call into the least expensive RAT in order to reduce the service cost incurred by the users. The service cost depends on the cost of the equipment and the cost of procuring spectrum license. This cost differs from one RAT to other RAT. It has the advantage of reducing overall service cost for the subscribers but it leads to high unbalanced network load.

F) Path loss based RAT selection algorithm

Path loss based RAT selection algorithm makes call admission algorithm based on path-loss measurements taken in the cells of each RAT. Path loss is carried out by measuring the received downlink power from a common control channel whose transmitted power is broadcast by the network. It has an advantage of low bit error rate and high throughput and it has the disadvantage of high frequency of vertical Handover. Perez Romero [13] proposes path-loss based RAT selection algorithm for initial RAT selection algorithm and Vertical Handover algorithm.

G) Network layer based RAT selection algorithm

Network layer based RAT selection algorithm admits calls into a particular layer. If the layer cannot accommodate the call, this algorithm tries to admit the call into the next layer. These algorithms are very simple but can lead to highly unbalanced network load. Network layer based RAT selection algorithm is explained in [14]. The objective of this algorithm is to minimize new call blocking probability while guaranteeing a hard constraint on handoff call dropping probability.

H) Utility/cost function based RAT selection

Incoming calls are admitted into a particular RAT based on some utility or cost function derived from a number of criteria. These algorithms are very efficient but are very complex and incur high computational overhead. [15] present the utility based RAT selection algorithms for selecting the RAT.

I) Mobile based RAT selection algorithm

This algorithm uses mobile terminal measurements from different radio access technology for the initial RAT selection [16]. The inputs to this algorithm are speed of the mobile user, signal strength, quality of service and service cost. This algorithm uses fuzzy logic controllers, genetic algorithms and particle swarm optimization for decision making under given input criteria. However, the mobile-based radio access technology selection algorithm requires higher computational power from mobile terminals.

J) Computational-intelligence based RAT Selection

Computational Intelligence based RAT selection admits an incoming call based on applying computational intelligence techniques for the call. The computational intelligence techniques applied for RAT selection are discussed

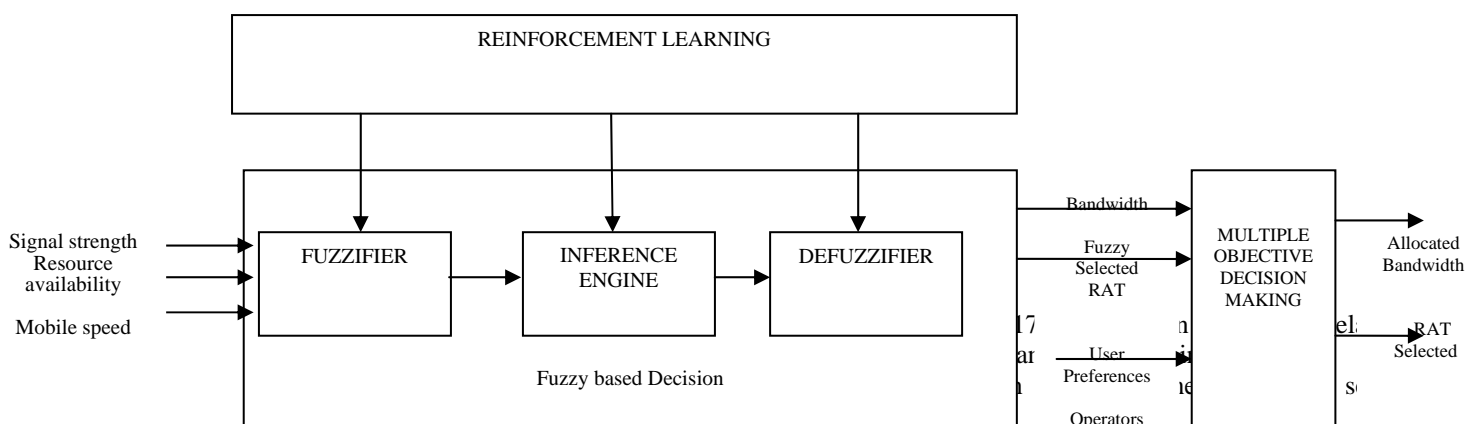


Fig. 3. Block Diagram of the Fuzzy Neural System

access network is integrated into an IP core network. This article proposed a new approach to handover management by applying a fuzzy logic concept to a heterogeneous environment. For handover initiation, parameters considered are network coverage, perceived QoS and Signal Strength (SS).

- The framework for JRRM algorithm [18-19] based on fuzzy neural mechanism as explained in Fig.3 consists of three main blocks namely fuzzy based decision, reinforcement learning and multiple objective decision making. The inputs for the fuzzy based decision block are signal strength of each RAT, resource availability of each RAT and mobile speed of the user.
- The Fuzzy based decision consists of three parts namely fuzzifier, inference engine and defuzzifier. The fuzzifier allocates a value from 0 to 1 for each input. In the inference engine, for each of the fuzzy subset defined in the fuzzifier, fuzzy rules are associated to indicate if it is suitable to be selected. The output of the inference engine is a value that varies between Y(yes), N(no), PY(probably yes) and PN(probably no). The defuzzifier will convert the output of the inference engine into fuzzy selected decision (FSD). The subjective and techno-economic criteria in the form of user preferences (UP) and operator preferences (OP) are inputs of the multiple objective decision making. The outputs of the fuzzy neural algorithm are cell/RAT selection and amount of bit rate allocated for the selected RAT.
- Fuzzy MADM (Multiple Attribute Decision Making) method [20-22] operates in two steps. The first step is to convert the imprecise fuzzy variables to crisp numbers. The second step is to use classical MADM technique to determine the ranking order of the candidate networks. The highest ranking RAT is then selected for the call.
- Using Fuzzy logic controllers, genetic algorithms and particle swarm optimization for decision making of radio access technology selection for the next generation wireless networks under given input criteria on user velocity, type of service and service parameters, Quality of service and service costs of the mobile user [23]. This algorithm uses mobile terminal measurements from different radio access technologies within a given time interval, with aim to obtain information for multi criteria decision making between different access networks available to the terminal.

VI CONCLUSION

In heterogeneous wireless networks, different RATs coexist in the same coverage area. The goal is to select the most suitable RAT for each user. The coexistence of different RATs requires a need for Joint Radio Resource Management (JRRM) to support the provision of quality of service and efficient utilization of radio resources. Hence this paper

presents the architecture for heterogeneous wireless networks and benefits for Joint radio resource management algorithms. Then an overview of Radio Access Technology selection for the Heterogeneous wireless networks is discussed. We analyze nine approaches for RAT selection among heterogeneous wireless networks and the advantages and disadvantages of each approach are also discussed. The future work in this area is to determine the best access technology among the available RATs by giving priority levels among the different classes of calls namely new calls, horizontal handoff calls and vertical handoff calls in heterogeneous wireless networks.

REFERENCES

- [1] *Radio Resource Management Strategies in UMTS*, J. Pérez-Romero, O. Sallent, R. Agustí, and M. A. Díaz-Guerra, Eds. Hoboken, NJ: Wiley, 2005.
- [2] V.K.Garg, *Wireless Communications and Networking*, Morgan Kaufmann Publishers, San Francisco, CA, USA 2007.
- [3] Marques, R. L. Aguiar, C. Garcia, J. I. Moreno, C. Beaujean, E. Melin, and M. Liebsch, "An IP-based QoS architecture for 4G operator scenarios," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 54–62, Jun. 2003.
- [4] J. Perez-Romero, O.Sallent, R. Agusti and M.A. Diaz-Guerra, *Radio Resource Management strategies in UMTS*, John Wiley & sons Ltd. Chichester, UK 2005.
- [5] D.Karabudakm C.Hung, B.Bing, "Call Admission control scheme using genetic algorithms", *Symposium on applied computing*, SAC'04, Nicosia, Cyprus, March 14-17, 2004.
- [6] Abdallah AL Sabbagh, Robin Braun, Mehran Abolhasan, "A Comprehensive Survey on RAT Selection Algorithms for Heterogeneous Networks", *World Academy of Science, Engineering and Technology* 73 2011.
- [7] O.E.Falowo and H.A.Chana, "Joint Call Admission control algorithms: Requirements, approaches and design considerations", *Computer Communications*, vol.31, no.6, April 2008, pp.1200-1217.
- [8] A.Tolli and P.Hakalin, "Adaptive Load Balancing between Multiple Cell layers", *2002 IEEE 56th Vehicular Technology Conference (VTC 2002)* vol.3., Vancouver, Canada, September 24-28, 2002, pp.1691-1695.
- [9] K.H.Suleiman, H.A.Chan, and M.E.Dlodlo, "Load Balancing in the Call Admission Control of Heterogeneous Wireless Networks", *International Conference on Communications and Mobile computing (IWCMC 2006)*, Vancouver, British Columbia, Canada, July 3-6, 2006, pp.245-250.
- [10] A.Umbert, L.Budzisz, N.Vucevic and F.Bernardo, "An all-IP heterogeneous wireless testbed for RAT selection and e2e QoS evaluation", *The 2007 International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 2007)*, Cardiff, Wales, UK, September 12-14, 2007, pp.310-315.
- [11] J.Perez-Romero, O.Sallent and R.Agusti, "Policy-based Initial RAT selection algorithms in Heterogeneous Networks", *The 7th IFIP International Conference on Mobile and Wireless Communications Networks (MWCN 2005)*, Marrakech, Morocco, September 19-21, 2005.
- [12] W.Zhang, "Performance of Real-time and Data traffic in Heterogeneous Overlay Wireless Networks", *19th International Teletraffic Congress (ITC 19)*, Beijing, China, August 29- September 2, 2005.
- [13] J.Perez-Romero, O.Sallent, R.Agusti, L.Wang, H.Aghavmi, "Network Controlled cell breathing for capacity improvement in heterogeneous CDMA/TDMA scenarios", *IEEE Wireless Communications and Networking conference (WCNC'06)*, April 3-6, Las Vegas, NV, USA.
- [14] R.B.Ali, S.Pierre, "An Efficient predictive admission control policy for heterogeneous wireless bandwidth allocation in next generation mobile networks", *International Conference on Communications and Mobile computing (IWCMC'06)*, Vancouver, Canada, July 3-6, 2006.
- [15] O.Ormond, J.Murphy, G.Muntean, "Utility-based Intelligent network selection in beyond 3G systems", *IEEE International Conference on communications (ICC 2006)*, Istanbul, Turkey, June 11-15.
- [16] Aleksandar Tudzarov and Toni Janevski, "Efficient Radio Access Technology Selection for the Next Generation Wireless Networks" in *International Journal of Research and Reviews in Next Generation Networks*, Vol. 1, No. 1, March 2011

- [17] P.M.L. Chan, R.E.Sheriff, Y.F.Hu, P.Conforto, C.Tocci, "Mobility management incorporating fuzzy logic for a heterogeneous IP environment", *IEEE Communications Magazine* 39(12) (2001) 42-51.
- [18] L. Giupponi, R. Agustí, J. Pérez-Romero, and O. Sallent, "A novel approach for joint radio resource management based on fuzzy neural methodology," in *IEEE Transactions on Vehicular Technology*, vol.57, No.3, May 2008.
- [19] R. Agustí, O. Sallent, J. Pérez-Romero, and L. Giupponi, "A fuzzyneural based approach for joint radio resource management in a beyond 3G framework," in *Proc. 1st Int. Conf. Quality Service Heterogeneous Wired/Wireless Netw.*, Dallas, TX, Oct. 2004, pp. 216–224.
- [20] W.Zhang, "Handover decision using fuzzy MADM in heterogeneous networks", *Proceedings of IEEE WCNC'04*, Atlanta, GA, March 2004.
- [21] A.L.Wilson, A.Lenghan, R.Malyan, "Optimizing wireless network selection to maintain QoS in heterogeneous wireless environments", *Proceeding of World Communication Forum*, Denmark, September, 2005.
- [22] O.E.Falowo and H.a.Anthony, "Fuzzy logic based call admission control for next generation wireless networks", *Proceedings of 3rd International symposium on wireless Communication Systems*, Valencia, Spain, September 5-8, 2006.
- [23] Aleksandar Tudzarov, Toni Janevski, "Efficient Radio Access Technology Selection for the Next Generation Wireless Networks", *International Journal for Research and Reviews in Next Generation Networks*, Vol.1, No.1, March 2011.

AUTHORS PROFILE



S. Palaniswami received the **B.E.** degree in electrical and electronics engineering from the Govt., college of Technology, Coimbatore, University of Madras, Madras, India, in 1981, the **M.E.** degree in electronics and communication engineering (Applied Electronics) from the Govt., college of Technology, Bharathiar University, Coimbatore, India, in 1986 and the **Ph.D.** degree in electrical engineering from the **PSG Technology**,

Coimbatore, India, in 2003. He was the Registrar of Anna University Coimbatore, Coimbatore, India, from May 2007 to May 2010. Currently he is heading the Department of Electrical and Electronics Engineering, His research interests include Control systems, Communication and Networks, Fuzzy logic and Networks, **AI**, Sensor Networks. . He has about 25 years of teaching experience, since 1982. He has served as lecturer, Associate Professor, Professor, Registrar and the life Member of **ISTE**, India.

J.Preethi received the **B.E.** degree in Computer Science and Engineering from Sri Ramakrishna Engineering College, Coimbatore, Anna University, Chennai, India, in 2003, the **M.E.** degree in Computer Science and Engineering from the Govt. college of Technology, Anna University, Chennai, India, in 2007 and she is currently pursuing the part time **Ph.D.** degree in the Department of Computer Science and Engineering from the Anna University Coimbatore, Coimbatore, India. Currently, she works as a Assistant Professor in the Department of Computer Science and Engineering, Anna University Coimbatore. Her research interests include Mobile adhoc networks, Mobile Communication systems especially in Radio Access Technology selection, Fuzzy logic and Neural Networks, Genetic Algorithms and **AI**.

Interactive Information System for online processing geo- technological data (GTD) sinking wells

Information Systems

Safarini Osama

IT Department
University of Tabuk,
Tabuk, KSA
usama.safarini@gmail.com
osafarini@ut.edu.sa

Abstract—: Online management of drilling requires the choice of an informed decision of many possible because of the volume of incoming and processed GTD, problem arising in the functioning through management situations. The importance here is the information management process to enable effective man-machine decision. So the purpose of work is to Develop a methodology, algorithm and program for processing (compression and classification) GTD sinking wells, confirming the geological GTD, for example, marks mining drill bits;

Keywords- Man-machine decision, Geo-technological data, classification, compression, correlation, measures of similarity, marks mining drill bits, data mining geology, Information Component, Euclidean and Hamming Distance.

I. INTRODUCTION

The *Work* describes methods and means of information system software for decision-making by the results of geo-technological data (GTD) on bore-hole drilling, compression of GTD on a drilling regime, improvement of interactivity of GTD processing and online management process of drilling, algorithms of GTD segmentation, a program product, results of GTD processing.

II. DISCUSSION

While classifying GTD, the process of segmentation aimed at taking On-line decisions in drilling, a forecast of the beginning or end of an interval, following and prediction of a working period of a drilling bit, evaluation of wear of drilling tools, prevention of emergency situations, breaking of equipment and others [1].

The results of the proposed segmentation provide us a geological situation through a well depth. The proposed methods assume interactive interpretation of segmentation and compression of GTD and a possibility of additional verifying repetitions and variations. This is connected with division of GTD into segments, their verification by the identified models applying two, essentially different methods:

- Separately for each well logging curve with their subsequent superposing for final segmentation;

- Obtaining of preliminary “integral” well logging curve and its segmentation.

From measures of similarity (see Table 1) is selected “distance indices” similar to a distance by Hamming and Euclid as the most widespread [2]. The features that describe distance indices in this case will be an amplitude and depth, while measures of similarity – their functions or as analogs of a distance by Hamming or Euclid:

- A product of a module of difference of amplitudes;
- A product of a module of amplitude difference by a difference of depths;
- A product of a module of amplitude difference by a square of differences of depths.

In these segmentation methods a number of segments, measures and functions applied here using the program shown in (Fig .1) can be varied with a possibility to present areas of segments, their models specifying borders, intersections, etc. very close to that which is now assumed for processing of vague sets as the measures of similarity of objects, classes are the values of the function that belongs to [3].

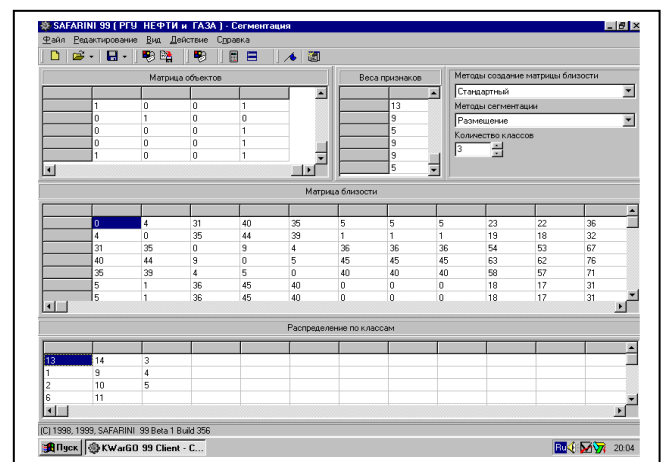


Fig.1 Program Interface for classification into classes

Classification by various measures of similarity Table 1			
Formula of a similarity measure	Division of information components		
	Class №1	Class №2	Class №3
1) $\frac{R_{ij}}{R_i + R_j - R_{ij}}$	1,3,4,5,6,7,8	11	2,9,10,12,13,14
2) $\frac{1}{r} \left[R_{ij} + \frac{R_{ij}}{r+1-(R_{ij}+R_{ij})} \right]$	5,3,4,6,7,8,9,11	14,13	2,1,10,12
3) $\frac{R_{ij}}{R_{ij} + R_{ij} + R_{ij}}$	1,3,4,5,6,7,8	11	2,9,10,12,13,14
4) $\frac{2R_{ij}}{2R_{ij} + R_{ij} + R_{ij}}$	1,3,4,5,6,7,8	11	2,9,10,12,13,14
5) $\frac{R_{ij} + R_{ij}}{r + R_{ij} + R_{ij}}$	5,3,4,6,7,8	1	14,2,9,10,11,12,13
6) $d_{ij} = \left(\sum_{k=1}^n C_k (x_{ik} - x_{jk})^2 \right)^{\frac{1}{2}}$ $C_{ij} = \frac{1}{1 + d_{ij}}$	5,3,4,6	14,9,10,11,13	1,2,7,8,12
7) $\sum_k x_{ik} - x_{jk} $	3,1,4,5	14,2,10,12	13,6,7,8,9,11
8) $\sum_k x_{ik} - x_{jk} $	3,1,2,4,5	14,9,10,11,12	13,6,7,8
9) $\frac{\sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sigma_i \sigma_j}$	5,4,6,7,8,9,11	14,2,10,12,13	3,1
10) $\frac{r-s}{r} \bar{C}_{ij} + \frac{s}{r} \bar{C}_{ij}$	5,3,4,6,7,8	14,2,9,10,11,12,13	1

Table 1. Classification by various measures of similarity

GTD Processing in two stages

- In the first stage compression and classification of GTD for each of the measurements. The results are a set of features for the second phase.
- At the second stage, the final classification on the full range of GTD, this allows assessing the correlation with marks of bits, or data mining geology.

Data Compression

Stage data compression involves the following steps:

- Calculate the autocorrelation function K_{xx} for every geo-measured properties curve.
- Determine the T_k - correlation interval for each sample. It is determined based on type of autocorrelation function.
- Approximation of each sample geo-measured properties sampling interval in depth equal to the T_k . In this case geo-measured data properties are

presented as a much simpler function with the same characteristics as the original sample.

- Classification of each geo-measured data properties.

1.000						
-0.514	1.000					
0.315						
-0.336	-0.135	1.000				
-0.498	-0.483					
-0.573	-0.117	0.651	1.000			
-0.554	-0.593	0.789				
0.435	0.288	-0.716	-0.924	1.000		
0.504	0.608	-0.746	-0.847			
0.627	0.175	-0.575	-0.805	0.848	1.000	
0.524	0.517	-0.473	-0.558	0.543		

Table 2. Correlation of parameters with / without separating into layers

	Average	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Layer 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Layer 2	0.289	-0.027	-0.585	-0.210	0.425	-0.243	-0.246
Layer 3	0.169	-0.037	0.171	0.093	-0.131	-0.195	-0.386
Layer 4	0.193	-0.030	-0.302	0.621	0.007	-0.045	0.152
Layer 5	0.274	-0.139	0.928	-0.109	0.416	0.000	-0.053
Layer 6	0.233	0.217	-0.911	0.015	-0.253	0.000	0.001
Layer 7	0.283	-0.348	0.911	-0.015	0.253	-0.165	-0.006
Layer 8	0.284	0.286	-0.827	-0.123	-0.374	-0.026	-0.066
Layer 9	0.262	-0.110	0.056	0.518	0.571	-0.313	-0.006
Layer 10	0.073	-0.043	-0.003	-0.043	0.017	0.075	0.256
Layer 11	0.123	0.036	-0.227	-0.362	0.005	-0.087	0.024
Layer 12	0.326	-0.806	0.791	0.069	0.064	0.071	0.158

Table 3. Correlation with the marks of mining bits

III. CONCLUSION

The developed information System is an instrument for decision-making in complicated multi-factor non-formalized cybernetic systems with a feedback, i.e.:

- in interactive assessment of informational significance of drilling factors provided by readings of on-land facilities, telemetric and feedback data [4];
- support of processing (compression and classification) of well sinking results verifying geological prospecting data, for instance, on a mark of drilling bit run;
- Developed are algorithms and programs for segmentation of GTD with a possibility of an interactive assessment of a segmentation quality,

variation of a number of segments, representativeness, correlation to a geological profile, borders of formations, wear of drilling bits, and prevention of emergency situations.

- Application of MS Excel for estimation of segments of GTD on a drilling regime;
- As seen from Table 2, the correlation shows better results when separating the well profile into layers; this reflects the fact of geology changing properties.
- As seen from Table 3, Correlation with the marks of mining bit confirms the changes in geo-measured data properties or as different layers.

REFERENCES

- [1] Levitzky A.Z., Komandrovsky V. G., Safarini Osama
Methods and Means to Develop an Information System for On-Line Control of Drilling, Scientific-Technical Research Journal, "Automation Telemetry and Communication in the Oil Industry", N 3 2000, PP 7-11.
- [2] Safarini Osama
"Enhanced Decision-Making Computer-Aided Methods for On-Line Control of Well Drilling", Abstracts of paper of the IPSI Conference Held in Carcassonne, France, UNESCO Heritage, April 27 to 30, 2006.
- [3] Levitzky A.Z., Komandrovsky V. G., Safarini Osama
On Automation for On-Line Control of Well Drilling, Scientific-Technical Research Journal, "Automation Telemetry and Communication in the Oil Industry", N 3-4 1999, PP 2-8.
- [4] Komandrovsky V. G., Safarini Osama
On classification of information components of On-Line control of a drilling Process, Abstract of paper of the Third Scientific Technical Conference, "Urgent Issues of the Condition and Development of the Oil and Gas Complex in Russia", Moscow, 1999 27-29 Jan.



AUTHOR'S PROFILE

Dr. **Safarini Osama Ahmad Salim** had finished his PhD. from The Russian State University of Oil and Gas Named after J. M. Gudkin, Moscow, 2000, at Computerized-Control Systems Department. He was awarded by his participation in Interpretation of measurement data in gas wells, Abstracts of paper presented to the third All- Russia Conference of young scientists, specialists and Students on the problems in gas industry in Russia "New technologies in the gas industry", Moscow 1999, 28-30 September.

He obtained his BSC and MSC in Engineering and Computing Science from Odessa Polytechnic National State University in Ukraine 1996. He worked in different countries and universities. His research is concentrated on Automation in different branches Specially Oil and Gaz.

Extended RR-scheduling algorithm

Prof. Sunita Chand, Ms. Teshu Chaudhary, Mr. Manoj Kumar

Krishna Engineering College,
95-Loni Road, Near Mohan Nagar, U.P-201007, India

e-mail: sunitamk@gmail.com,

teshuchaudhary@gmail.com

kumarmanoj1989.kumar@gmail.com.

Abstract: RR-scheduling algorithm was designed for the time-sharing system or interactive systems. The first process in the queue run until it expires its quantum (i.e. runs for as long as the time quantum), then the next process in the queue runs and so on. RR scheduling is implemented with timer interrupts. When a process is scheduled, the timer is set to go off after the time quantum amount of time expires. When process expire its quantum, a context switch takes place. The state of the running process is saved and context of next process in the ready queue is loaded in CPU registers. it gives good response time, but can give bad waiting time.

We propose here a modification to round robin scheduling algorithm which not only gives good response time but also shows reduction in waiting time. If the processes in the ready queue are arranged in the increasing order of the expected CPU burst time instead of first come first serve manner, the waiting time of the processes will decrease in addition to fast response time.

I. INTRODUCTION:

In this approach, the ready queue is assumed to be a circular queue. Extended RR-scheduling algorithm is also designed for the time-sharing system. This algorithm may prove to be better than RR-scheduling algorithm in following ways:

- It reduces waiting time
- It reduces turn around time
- It reduces response time
- In some case, context switching time can be reduced.
- If two or more process has same burst time then a process that has highest priority will get the CPU first. The highest priority process will has to no longer wait in ready queue.

II. EXTENDED ROUND ROBIN SCHEDULING ALGORITHM

The Extended round robin algorithm works in the following way :

- i. The processes are evaluated on the grounds of expected CPU burst time and are arranged in the ready queue in the increasing order of CPU time
- ii. The ready queue is maintained as a circular queue.
- iii. The processes may be considered to arrive at the same time. In such a case arrival time for all process is considered to be zero.
- iv. There may be cases when arrival time of the processes are different. In that case the ready queue needs to be refreshed every time a new process arrives in the system according to the shortest CPU burst time of all the processes in ready queue along with the newly entered process.
- v. No process can hold the CPU forever. Each process executes for the a period of time slice.
- vi. Time Sharing is implemented by a hardware timer. On each context switch, the system loads the timer with the duration of time slice and hands control over to the new process. The preempted process is re- queued at the end of the ready queue. When the timer times out, it interrupts the CPU which then steps in and switches to the next process.
- vii. Concept of priority is used to resolve the contentions that may result when two or more processes have the same burst time (execution time), in that case the CPU is allocated to that process which needed the CPU quickly and want to finish in short time or we can say has higher priority.
- viii. When short processes keep entering in system, long process will suffer starvation as every time a short process enters the system, the ready queue will be refreshed and the longer process will be shifted to the tail of ready queue. Although starvation cannot be removed completely, it can be minimized by using AGING. Whenever a process is put to the tail of ready queue without execution, the priority of this process should be increased by one i.e., numerically it should be decreased as low number represent high priority. This way the process will get priority and gets a chance to execute.

III. EVALUATION OF EXTENDED RR-SCHEDULING ALGORITHM:

We evaluate the new “Extended RR Scheduling Algorithm” using deterministic modeling approach of “analytic evaluation” method [1]. It takes a particular pre-defined workload and evaluate the performance of each algorithm for that workload. Here we determine the behavior of normal round robin and extended round robin algorithm presented here, for the same work load. Also we check the response of these algorithms for different set of workload. We consider them as different cases. Each case is explained separately to compare the performances of the two algorithms.

Also the performance of the Extended RR scheduling has been proved to be far better as compared to normal RR Scheduling algorithm through the C-Code implementation. The code written in C language executed for a variety of sets of workload for different number of processes, thus proving the above fact .

CASE 1:

IF an arrival time of all the processes is assumed to be same and their burst times are different.

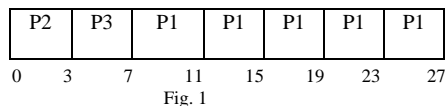
THEN Extended RR Scheduling algorithm sort the processes in increasing order and allocate these processes to CPU in the same order. (No need of priority in this case)

EXAMPLE 1:

Process	Burst time
P1	20
P2	3
P3	4

Arrival time for each process is 0. Time slice is 4 ms.

Gantt Chart for extended RR Scheduling is as shown in Fig.1:



Waiting time:

P1 : 7

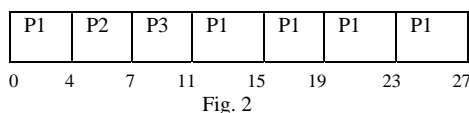
P2 : 0

P3 : 3

Average Waiting time: $(7+0+3)/3 = 10/3$;

SOLVED BY RR-SCHEDULING ALGORITHM :

Gantt Chart for RR Scheduling is as shown below in Fig. 2.



(According to RR scheduling Average Waiting time is $= (7+4+7)/3 = 18/3$)

NOTE: In some case context-switching time can be reducing.

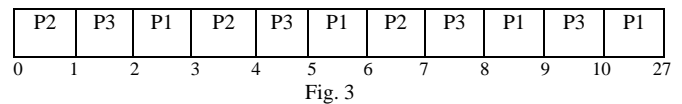
In the above solution Context switch takes place 4 times while in the previous solution context switch took place only 3 times

EXAMPLE 2:

For the same set of processes, consider that the arrival time for each process is 0, whereas Time slice is 1 ms.

Figure 3 below shows the behavior of “Extended RR scheduling algorithm” whereas figure 4 shows the behavior of normal RR algorithm.

SOLUTION USING EXTENDED RR SCHEDULING ALGORITHM:



Waiting time:

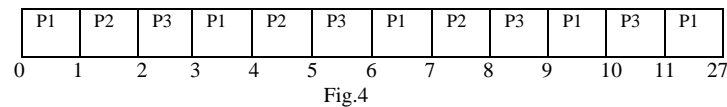
P1 : $2+2+2+1 = 6$

P2 : $0+2+2 = 4$

P3 : $1+2+2+1 = 6$

Average Waiting time: $(6+4+6)/3 = 18/3$;

Solution using Normal RR Scheduling Algorithm:



Waiting time:

P1 : $0+2+2+2+1 = 7$

P2 : $1+2+2 = 5$

P3 : $2+2+2+1 = 7$

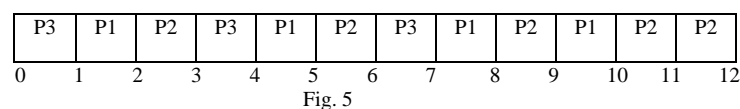
(According to RR scheduling Average Waiting time is $= (7+5+7)/3 = 19/3$;

Example 3:

Process	Burst time
P1	4
P2	5
P3	3

Arrival time for each process is 0. Time slice is 1 ms.

Gantt Chart for Extended RR Scheduling is as shown in Fig.5:



Waiting time:

P1 : $1+2+2+1 = 6$

P2 : $2+2+2+1 = 7$

$$P3 : 0+2+2 = 4$$

Average Waiting time: $(6+7+4)/3 = 17/3$;
(According to RR scheduling Average Waiting time is = $19/3$)

CASE 2:

IF an arrival time of all the processes is the same and the burst time of some of these processes are also same.
THEN sort the processes in increasing order according to the burst time and sort those processes having same burst time according to their priority (highest priority process will get the CPU first). we compare a new process to all process excluding lastly executed process(example 2).

EXAMPLE 1:

Process	Burst time	Priority
P1	20	3
P2	3	2
P3	3	1

Arrival time for each process is 0. Time slice is 4 ms.

Gantt Chart for extended RR Scheduling is as shown in Fig.6:

P3	P2	P1	P1	P1	P1	P1	
0	3	6	10	14	18	22	26

Fig. 6

Waiting time:

P1 : 6
P2 : 3
P3 : 0

Average Waiting time: $(6+3+0)/3 = 9/3$;
(According to RR scheduling Average Waiting time is = $17/3$).

EXAMPLE 2 :

Process	Burst time	Priority
P1	6	
P2	4	2
P3	3	
P4	4	1

Arrival time for each process is 0. Time slice is 3 ms.

Gantt Chart for extended RR Scheduling is as shown in Fig.7:

P3	P4	P2	P4	P2	P1	P1	
0	3	6	9	10	11	14	17

Fig. 7

Waiting time:

P1 : $9+2 = 11$
P2 : $6+4 = 10$
P3 : $0 = 0$
P4 : $3+6 = 6$

Average Waiting time using extended RR Scheduling : $(11+7+0+6)/4 = 24/4$.

(According to normal RR Scheduling, Average Waiting time is = $40/4$).

CASE 3

IF an arrival time of all the processes is different i.e., processes in the system are arriving at different time which is quite obvious, and the CPU burst time of some of them are same.

THEN the processes need to be sorted in increasing order of the CPU burst time every time when a new process arrives in the system. and the processes having same burst time are sorted according to their priority (highest priority process will get the CPU first).The CPU is allocated to these processes in the round robin manner for a time period equal to the “time slice” or “time quantum”.

We describe this algorithm as below:

IF the time of process completion (or partial completion) is equal to the time of a new process entering into the queue.

THEN compare burst time of new process to the remaining burst time of processes **excluding last executed process**.

IF burst time of new process is less than the remaining processes

THEN firstly, new process should be allocated to CPU then other remaining processes

ELSE sorts all remaining process as per their remaining burst time together with the new process and then allocate the CPU to these processes in increasing order of their CPU burst time.

EXAMPLE 1.

Process	Arrival time	Burst time	Priority
P1	0	10	1
P2	1	5	3
P3	2	5	2

Time slice is 3 ms.

Gantt Chart for Extended RR Scheduling is as shown in Fig.8:

P1	P3	P2	P3	P2	P1	P1	P1	
0	3	6	9	11	13	16	19	20

Fig.8

Waiting time:

$$\begin{aligned} P1 : 0+10-0 &= 10 \\ P2 : 6+2-1 &= 7 \\ P3 : 3+3-2 &= 4 \end{aligned}$$

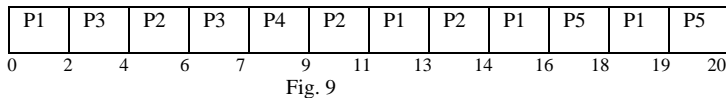
Average Waiting time: $(10+7+4)/3 = 21/3$;
(According to RR scheduling Average Waiting time is = 27/3.

EXAMPLE 2

Process	Arrival time	Burst time	Priority
P1	0	7	1
P2	1	5	2
P3	2	3	4
P4	6	2	3
P5	12	3	5

Time slice is 2 ms.

Gantt Chart for Extended RR Scheduling is as shown in Fig.9:

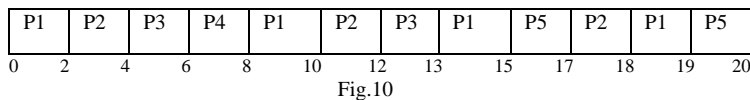


Waiting time:

$$\begin{aligned} P1 : 0+9+1+2-0 &= 12 \\ P2 : 4+3+2-1 &= 8 \\ P3 : 2+2-2 &= 2 \\ P4 : 7-6 &= 1 \\ P5 : 16+1-12 &= 5 \end{aligned}$$

Average Waiting time as per Extended RR scheduling algorithm : $(12+8+2+1+5)/5 = 28/5$

Gantt Chart for normal RR Scheduling is as shown in Fig.10:



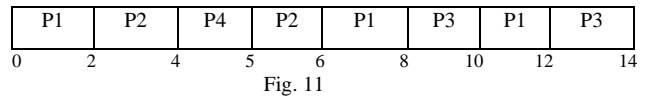
Average Waiting Time as per normal Round Robin Scheduling Algorithm : $(12+12+8+0+5)/5 = 37/5$

EXAMPLE 3:

Process	Arrival time	Burst time	Priority
P1	0	6	1
P2	1	3	3
P3	2	4	4
P4	4	1	2

Time slice is 2 ms.

Gantt Chart for Extended RR Scheduling is as shown in Fig.11:



Waiting time:

$$\begin{aligned} P1 : 0+4+2-0 &= 6 \\ P2 : 2+1-1 &= 2 \\ P3 : 8+2-2 &= 8 \\ P4 : 0-4 &= 0 \end{aligned}$$

Average Waiting time as per Extended Round Robin Scheduling : $(6+2+8+0)/4 = 16/4$

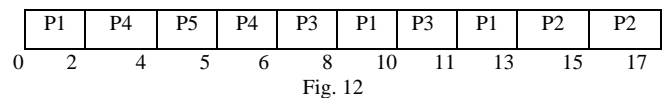
Average Waiting time as per normal Round Robin Scheduling = 22/4

EXAMPLE 4:

Process	Arrival time	Burst time	Priority
P1	0	6	1
P2	1	4	2
P3	2	3	4
P4	2	3	3
P5	4	1	5

Time slice is 2 ms.

Gantt Chart for Extended RR Scheduling is as shown in Fig.12:



Waiting time:

$$\begin{aligned} P1 : 0+6+1-0 &= 7 \\ P2 : 13-1 &= 12 \\ P3 : 6+2-2 &= 6 \\ P4 : 2+1-2 &= 1 \\ P5 : 4-4 &= 0 \end{aligned}$$

Average Waiting time using Extended RR scheduling: $(7+12+6+1+0)/5 = 26/5$;

Average Waiting time using normal RR scheduling: 38/4.

IV. CONCLUSION AND FUTURE SCOPE

Extended RR-scheduling algorithm can reduce the Waiting time, turnaround time and the response time. If two or more processes have same burst time then a process that has higher priority will get the CPU. The time sharing system can become more effective from the point of view of Average Waiting Time and Turnaround time.

Although there are chances of longer processes to be starved when shorter processes keep entering the system. In that case aging may prove to be helpful in providing the solution.

There can be other solutions to this problem like implementing a separate queue for those longer process which reaches at the head of ready queue for execution but CPU is not allocated to these process as some other short process has entered the ready queue. We are leaving this solution to be evaluated as our future scope.

V. REFERENCES

- [1]. *Operating System Concept*, Silberschatz Galvin.

Enhancement of Throughput for Single Hop WPAN's using UWB- OFDM Physical Layer

Ch. Subrahmanyam
Department of ECE
Scient Institute of Technology
Hyderabad, India
Email: subbunvl@yahoo.com

K. Chennakesava Reddy
Department of ECE
TKR College of Engg. & Tech.
Hyderabad, India
Email: kesavary@hotmail.com

Syed Abdul Sattar
Department of ECE
Royal Institute of Tech. & Science
Hyderabad, India
Email: syedabdulsattar1965@gmail.com

Abstract— One of the significant and secure agents for the UWB (Ultra Wide Band) based alternative physical layer for WPAN's (Wireless Personal Area Networks) is MB – OFDM (Multiband Orthogonal Frequency Division Multiplexing). In this presentation, the simulation ejaculates for single Hop WPAN depending upon the OFDM UWB physical layer are expounded. In this effect, the transmittal systems for the data progressions of 55 Mbps, 200 Mbps, and 480 Mbps are applied because these three are correspondents for lowest progression, the highest the compulsion rate and the greatest optional rate resultantly. We applied both 4mX 4m and 10mX10m insular fields for the network regions for the single Hop sketches in the simulation designs. The prevalent functions of the single Hop WPANS like average End – to – End Delay and Packet Failure Rate(PFR) and Throughputs for the entire source – target oriented pairs are replicated by imparting the Qualnet network simulator.

Keywords- OFDM, Single Hop, Throughput, UWB, WPAN's

I. INTRODUCTION

Nowadays, we have the requirement for wireless communication systems which could be manipulated at a huge amount of data progressions over short distance communications so as to meet the sophisticated product outcomes in consumer electronics i.e. Camcorders, DVD Players, etc. The utmost utilization of high-end Wireless Personal Area Networks (WPANS) for short distances with improvised connectivity among consumable electronics and interactive devices have got established more prominently since 2000. Having been approved by the Federal Communications Commission (FCC) for the application of Ultra- Wide- Band (UWB) on the unlicensed band in 3.1 – 10.6 GHz range, this enhances the extensive usage of high speed WPAN systems (up to 480 Mbps) standing on a UWB physical layer execution. The renowned IEEE 802.15.3 has been structured with the same high- rate WPANS by the special interest group (SIG).

In this methodological script, at first we begin with a comprehensive conception of a Wireless Personal Area Network (WPAN), next introduction of fundamentals for

UWB radio communications and later the presentation on concept of single Hop WPAN's in brief. Finally in the simulations, major limitations of single Hop adhoc WPAN's can be discussed.

A. Overview of WPAN's

Wireless Personal Area Networks (WPANs) capacitate the lower distant expedient connectivity among compact consumer electronics and communication devices. The range of a WPAN is generally restricted to a radius of 10 meters. The Bluetooth radio system has materialized as the first electronic component representing WPAN applications with its prominent elements of low power consumption, small in size, and low in cost. Data weight for Bluetooth devices is restricted to 1 Mbps for version 1.2, and 3 Mbps for version 2.0 with enhanced data rate (EDR), respectively. These data tariffs are adequate for streaming stereo's audio, transmitting data or carrying voice communications, but they are not sufficient to back up for multimedia traffic. The IEEE 802.15.1 Standard was extracted from the Bluetooth version 1.1 Foundation Specifications, and was published in June 2002.

The next generation of consumer oriented compact electronics and communications devices will support multimedia data traffic that requires high data rates. These applications contain high-quality video and audio distribution, multi-megabyte file transmissions for music and image files [1]. For example, devices that will use high-rate WPANs include digital camcorders, digital televisions, digital cameras, MP3 players, printers, projectors, and laptops, etc [1]. The requirement for communications between these multimedia-capable devices leads to associated judicious type connections that warrant data rates well in excess of 20 Mbps and Quality of Service (QoS) provisions with respect to guaranteed bandwidth [1]. To assimilate the required physical layer and MAC layer QoS requirements, the IEEE 802.15 WPAN Working Group initiated a new group i.e. the 802.15.3 High-Rate WPAN Task Group. The IEEE 802.15.3 Standard was framed to capacitate wireless connectivity of high-speed, low-power, low-cost, multimedia-capable consumer electronic devices [10]. The idea of adding high-rate strength to the IEEE 802.15 family of standards was first incorporated in November 1999. The

802.15.3 Task Group started their official work in March 2000, and 802.15.3 was finally accepted as an IEEE Standard in June 2003. This Standard is not expected to be a plain enlargement of the IEEE 802.15.1 Standard because the MAC needs is very variant.

Conventionally, an IEEE 802.15.3 compliant WPAN engages in an unlicensed 2.4 GHz frequency range with an RF bandwidth of 15 MHz. The symbol progression is 11 Mbps and applies to all specified modulation formats, including QPSK, DQPSK, and 16/32/64 QAM [1]. Through the use of multi-bit symbol modulation and channel coding, the attainable data rates can be in the amplitude from 11 Mbps to 55 Mbps. a much higher data rate is required than that specified in the IEEE 802.15.3 Standard, for applications that involve imaging and multimedia, such as H.323/T.120 video conferences, home theatre, interactive applications, and file downloading. To enumerate a project to facilitate a higher speed PHY enhancement correction to 802.15.3 for these applications, the IEEE 802.15 High Rate Alternative PHY Task Group (TG3a) for WPANs was constituted. This alternative physical layer (alt-PHY) is intended as a supplement to the IEEE 802.15.3 range. To be supported by the physical layer, a bit rate of at least 110 Mb/s at a distance of 10 meters is required. The transmission strength is ensured static by supervisory emission limits. An accumulating higher bit rate of at least 200 Mb/s at a distance of 4 meters is required. Even at the expense of reduced operating distances, scalability to rates in excess of 480 Mb/s is expected. The Data rates in the actual proposals may be higher, data rates mentioned above are minimums and most proposals favor the Ultra Wide Band physical layer implementation approach to realize the desired system specifications.

To dispatch information over comparably lowest destinations among a few participants [10], Wireless personal area networks (WPANs) are utilized. A WPAN is distinguished from other types of data grids. In that, communications are normally decentralized to a minute area that literally covers about 10 meters in radius and totally covers connected equipment whether static or in motion. High-Rate WPAN activates multimedia relation among compact instruments within a Personal Operating Space (POS). A set of devices within a POS, which control under the control of a Pico net controller (PNC) in order to share a wireless resource, is called a Pico net. The basic timing for the WPAN is to offer the function of the PNC. Additionally, the PNC manages the Quality-of-Service (QoS) requirements for the WPAN as a whole.

B. UWB radio Communications-Its Fundamentals

The IEEE 802.15.3 High Rate Alternative PHY Task Group (TG3a) for WPANs is functioning is to ascertain a project to facilitate a higher speed PHY enhancement amendment to 802.15.3 in order to support very high data rate applications as

mentioned in the previous section., The goals for this standard are to attain and obtain data rates of up to 110 Mbps at a 10 m distance, 200 Mbps at a 4 m distance, and higher data rate at shorter distances [7]. Depending upon these criteria, various proposals were acceded in response to 802.15.3a. Most proposals favor the Ultra-Wide-Band (UWB) physical layer. UWB systems have shown their ability to satisfy such needs by providing data rates of up to several hundred Mbps. UWB was first used to directly modulate an impulse-like waveform with very short duration occupying several gigahertz of bandwidth. Two examples of such systems are Time-Hopping Pulse Position Modulation (TH-PPM) and Direct-Sequence UWB (DS-UWB). Imparting these conventional UWB methods over the entire allocated frequency, band has many disadvantages, including need for high complexity RAKE receivers to capture multipath energy, high-speed analog-to-digital converters (ADC) and high power consumption. These considerations motivated a shift in the UWB system design method from initial "Single-Band" radio that occupied the entire allocated spectrum in favor of a "Multi-Band" design strategy [2]. According to the FCC "Multi-Band" schemes divide the available UWB spectrum into several sub-bands, each one occupying approximately 500 MHz (which is the minimum bandwidth for a UWB system definition). As if it were following the total of its bandwidth by interleaving symbols across different sub-bands, a UWB system can still organize the same transmit power. A narrower sub-band bandwidth also calms down the necessity on the sampling rate for ADCs consequently enhancing digital processing capability [2]. Multiband-OFDM (MB-OFDM) is one of the promising candidates for the alternative PHY layer implementation to facilitate WPANs. It combines Orthogonal Frequency Division Multiplexing (OFDM) with the above described multi-band method activating UWB transmission in order to inherit all the strengths of an OFDM technique which has already proven its unique role in wireless communications systems (ADSL, DVB, 802.11a, 802.16.a, etc) [2].

II. SINGLE HOP WPAN'S

In this section, based on the OFDM UWB physical layer are presented, the simulation yields for single-hop WPAN. The objective in using a single-hop scenario is to evaluate the Physical and MAC developed in this analysis. Transmission systems for rates of 55 Mbps, 200 Mbps, and 480 Mbps are interpolated in this observation as they are representative of the lowest rate, the highest mandatory rate and the highest optional rate, respectively. Both the 4m x 4m and the 10m x 10m circuitry areas for the network regions are used for simulation studies of the single-hop scenarios. Since the transmission radii of MBOA OFDM UWB systems that achieve a PER of 8% are 12.0 m, 7.4 m, and 3.2 m for systems organizing at 55 Mbps, 200 Mbps and 480 Mbps, respectively, the progressive functioning of the single-hop WPAN is easily apprehended within these network areas. In addition to the average end-to-end delay and packet failure rate, the total throughputs for all source- destination pairs are also gained.

Table 1. analyses the system limitations used in the simulations for the single-hop scenarios considered in this study.

Simulation parameter	Value
Simulation Time	1s
Number of nodes	20
Number of links	2,4,6,8,10
Network Area	4mX4m, 10mX10m
Number of Channels	1(Center Frequency = 3.432 GHz)
Transmission Power	-10.3 dBm
Receiver sensitivity	-77.2 dBm for 200 Mbps -72.6 dBm for 480 dBm
Channel model considered	Free space,Shadowing,an Rayleigh fading
Packet size(application layer)	982 bytes(will be 1024 bytes after MAC layer)
Max Network Buffer size	5,000 Bytes
Number of source Destination pairs	2,4,6,8,10
Guard time between slots	1 μ s
Intra Frame time	1.875 μ s

The packet loss because of collisions will be negligible since the number of slots per model is set to be the number of source-destination pairs, and each active source node is assigned one time slot within one frame,. The efficiency for scheduling should be close to 100%, theoretically. Generally, when the network saturation is reached, the packet failure rate will be increased dramatically due to the buffer overflow, and the average delay will also be enhanced due to extensive marking.

A. 4m X 4m Single Hop System

The average delay, PFR, and throughput functioning for the single-hop condition within the $4\text{m} \times 4\text{m}$ network area are specified in Figures 1 to 3 as a function of the number of source-destination pairs. Since the base and targeting nodes are haphazardly applied, the average distance for the active links will be less than 3 m. It is elicited from Section II.A that the propagation ranges to obtain 8% PER for systems operating at 55 Mbps, 200 Mbps and 480 Mbps are about 12 m, 7.4 m, and 3.2 m, respectively. Therefore, if the physical layer and MAC layer are developed in this research work perfectly (that is, the system performance match those presented on the MBOA proposal, when considering the overheads of other network layers), the packet failure rate due to the blunder in channel will be very minute (close to zero) for transmission systems operating at 55 Mbps and 200 Mbps. However, there may be transmission errors present for transmission systems operating at 480 Mbps.

The saturation throughput, that is attainable throughput when the network saturation occurs for transmission systems operating at 55 Mbps, is reached when 8 or more source-destination pairs are present. This is appropriate since the network throughput produced in [8] is about 48 Mbps at the physical layer for systems being managed at 55 Mbps. It can be noticed that the average delay is less than 5 ms, and the

PFR is close to zero before the occurrence of throughput saturation. After saturation throughput (about 44 Mbps) has been reached, the average delay is rated to over 60 ms, and the PFR is increased to over 8%. It can be summarized that for single-hop scenarios within a $4\text{m} \times 4\text{m}$ area, the performance measures for average delay and PFR are both feasible and meet QoS needs before the congestion of throughput is arrived for transmission systems operating at 55 Mbps supporting real-time applications. it can be observed that the network saturations are not reached even when 10 source-destination pairs are present for transmission systems operating at both 200 Mbps and 480 Mbps,. These results are appropriate since the network throughputs presented in [8] are about 120 Mbps and 180 Mbps for transmission systems operating at 200 Mbps and 480 Mbps. Both the average delay ($<5\text{ms}$) and PFR ($<5\%$) are small in this case. The PFR is slender enhanced (from 0.072% to 1.38%) collated to that for the 200 Mbps transmission for machines manipulating at 480 Mbps, however, it is bounded in the feasible range ($<5\%$). It has been observed that the simulation yields for $4\text{m} \times 4\text{m}$ single-hop situations equate those produced in the MBOA OFDM UWB proposal, and the physical layer and MAC layer improvised in this study work well for $4\text{m} \times 4\text{m}$ single-hop correspondence system architecture.

B. Simulation Results for 4m X4m Single Hop System

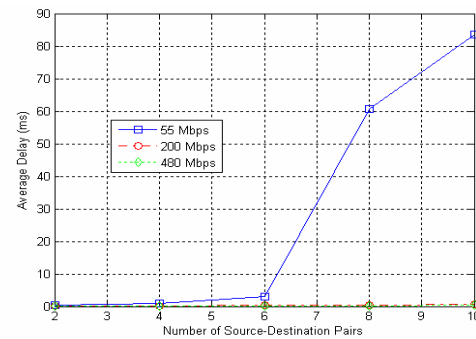


Figure 1.: Average End-to-End Delay vs. Number of Source-Destination Pairs for Single-Hop:4mX4m Area

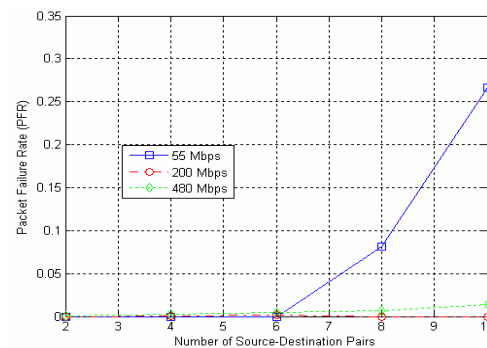


Figure 2.: PFR vs. Number of Source - Destination pairs for the single Hop scenario: 4mX4m Area.

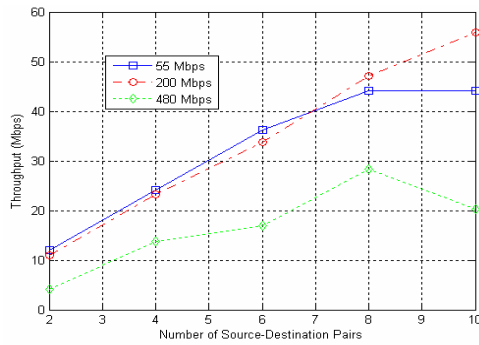


Figure 3.: Throughput vs. Number of Source – Destination pairs for single- Hop scenario: 4mx4m Area

C. Simulation Results for 10m X 10m Single Hop System

The average delay, PFR, and throughput performance for the single-hop scenarios within a 10m x 10m area are illustrated in Figures 4 to 6 as a function of the number of source-destination pairs. Since the source and destination nodes are randomly assigned, the average distance for the active links will be less than 7 m. Theoretically, if the physical layer and MAC layer developed in this study work well, the packet failure rate due to the channel error will be very small (close to zero) for the systems operating at 55 Mbps. However, there may be channel errors present for the systems operating at 200 Mbps. It was estimated that there would be a huge number of channel errors available within those systems functioning at 480 Mbps.

It can be analyzed that for systems manipulating at 55 Mbps, since it is still within the propagation range (about 12 m) in this case, the execution is approximately similar to the 4m x 4m area. For systems operating at 200 Mbps, the due point productivity is not reached even when 10 source-destination pairs are present. The average downtime is very minimal, and less than 10 ms. The PFR is between 4% and 8%, which is much greater than the PFR gained in the case of a 4m x 4m geographical network area. The saturation productivity is not approached at least for 10 source-destination pairs, for systems operating at 480 Mbps. The average downtime is very minimal, and less than 10 ms. However, the PFR is between 40% and 70%, which is too large to be acceptable. It can be seen that the achievable productivity for machines ordination at 480 Mbps is much less than those for systems operating at 55 Mbps and 200 Mbps. This is because more packets are dropped due to the presence of higher channel BER. It has been extracted and understood that the simulation outcomes for 10m x 10m single-hop scenarios equate those produced in MBOA OFDM UWB proposition, and the physical layer and MAC layer empowered in this study execute well for a 10m x 10m single-hop communication system configuration.

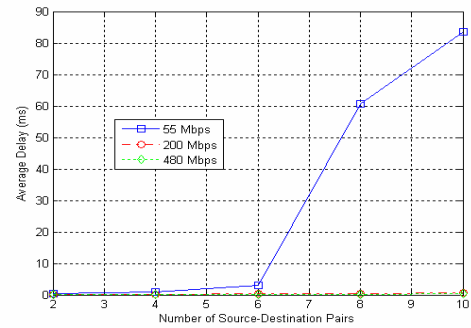


Figure 4.: Average End-to-End Delay vs. Number of Source-Destination Pairs for single Hop scenario: 10mX10m Area

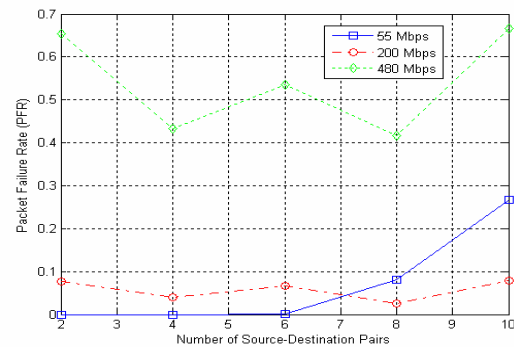


Figure 5.: PFR vs. Number of Source Destination pairs for the single Hop scenario: 10mX10m Area

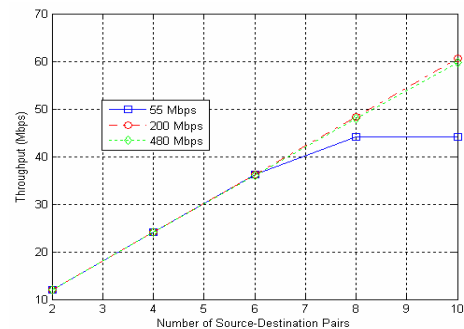


Figure 6.: Throughput vs. Number of Source – Destination pairs for single Hop scenario: 10mX10m Area

III CONCLUSIONS

Considering the results for both the 4m x 4m and 10m x 10m geographical network regions, it can be verified that for single-hop WPAN systems, within the coverage radius, before the saturation throughput is reached, the criteria of performance for all data rates or progressions (55, 200 and 480 Mbps), i.e. the average delay or downtime and PFR, arrive at the QoS requirements for real-time applications.

REFERENCES

- [1] J. Karaoguz, "High-Rate Wireless Personal Area Networks," IEEE Communications Magazine, Vol. 39, pp. 96-102, December 2001.
- [2] S. M. S. Sadough, A. Mahmood, E. Jaffrot and P. Duhamel, "Performance Evaluation of IEEE 802.15.3a Physical Layer Proposal Based on Multiband- OFDM," <http://www.lss.superlec.fr/>.
- [3] R. Bruno, M. Conti and E. Gregori, "Mesh Networks: Commodity Multihop Ad Hoc Networks," IEEE Communications Magazine, pp. 123-131, March 2005.
- [4] C. S. Murthy and B. S. Manoj, Ad Hoc Wireless Networks: Architecture and Protocols, Prentice-Hall, NJ, 2004.
- [5] A. F. Molisch, J. R. Foerster and M. Pendergrass, "Channel Models for Ultra wideband Personal Area Network," IEEE Wireless Communications Magazine, Vol. 10, pp. 14-21, December 2003.
- [6] M. D. Benedetto and G. Giancola, Understanding Ultra Wide Band Radio Fundamentals, Prentice-Hall, NJ, 2004.
- [7] L. Maret, I. Siaud and Y. Kamiya, "Ultra Wideband PHY Layer MBOA Performance and Sensitivity to Multipath Channels (IST Magnet Project)," <http://www.ist-magnet.org/>.
- [8] Multiband OFDM Alliance, "Multi-Band OFDM Physical Layer Proposal for IEEE 802.15 Task Group 3a," September 14, 2004, <http://www.wimedia.org/>.
- [9] H. Xu and A. Ganz, "A Radio Resource Control Method in UWB Protocol Design
- [10] P. Mohapatra, J. Li and C. Gui, "QoS in Mobile Ad Hoc Networks," IEEE Wireless Communications Magazine, pp. 44-52, June 2003.

AUTHORS PROFILE



Prof. Ch. Subrahmanyam presently working as a Professor & Head, Department of ECE at Scient Institute of technology, Hyderabad. He has completed his B.E. in 1995 from Andhra University, A. P. India, and M. Tech. from JNTU Hyderabad, in 2002, and Pursuing his Ph.D. from JNTU Hyderabad, A. P. India with ECE in Wireless communications. He has about 15 years of experience in teaching and industry together, he is having publications in International Journals and Conferences. He has guided many M. Tech and B. Tech. Projects. He is a life member of ISTE, India.



Dr. K. Chennakeshava Reddy, Presently working as Principal & Professor of ECE at TKR College of Engineering. He has completed his B.E. in 1973 and M. Tech. in 1976 from REC Warangal, A.P. India, and Ph.D. in 2001 from JNTU Hyderabad. He has worked in various positions starting from lecturer to Director of Evaluation in JNT University, Hyderabad, A. P. India. He has about 70 publications in international and National journals and Conferences and he has successfully guided 4 Ph.Ds and many are under progress. He is a member of various technical Associations.



Dr. Syed Abdul Sattar, presently working as a Dean of Academics & Professor of ECE department, RITS, Chevella, Hyderabad. He has completed his B.E. in ECE in 1990 from Marathwada university Aurangabad, M.S. India, M. Tech. In DSCE from JNTU Hyderabad, in 2002, and Pursued his first Ph.D. from Golden state University USA, with Computer Science in 2004, and second Ph.D. from JNTU Hyderabad, A. P. India with ECE in 2007. His area of specialization is wireless communications and image Processing. He has about 21 years of experience in teaching and industry together and recipient of national award as an Engineering Scientist of the year 2006 by NESI New Delhi, India. He has about 73 publications in International and National Journals and conferences. Presently he is guiding more than 15 research scholars in ECE and Computer Science from different Universities. He is a member of Board of studies for a central university and reviewer/editorial member/chief editor for national and International journals.

Enhancement of Throughput for Multi Hop WPAN's using UWB- OFDM Physical Layer

Ch. Subrahmanyam
Department of ECE
Scient Institute of Technology
Hyderabad, India
e-mail: subbunvl@yahoo.com

K. Chennakesava Reddy
Department of ECE
TKR College of Engg. &Tech.
Hyderabad, India
e-mail: kesavary@hotmail.com

Syed Abdul Sattar
Department of ECE
Royal Institute of Tech. & Science
Hyderabad, India
Email: syedabdulsattar1965@gmail.com

Abstract— One of the most significant determinants for the UWB (Ultra Wide Band) based substitutive physical layer for WPANS (Wireless Personal Area Networks) is MB – OFDM (Multiband Orthogonal Frequency Division Multiplexing). This paper deals in the manipulation outcomes for Multi-Hop WPAN depending upon the UWB - OFDM physical layer are exhibited. However, the spectrum radius of MB-OFDM UWB machines is quite minimal, and single-hop transmissions may not be sufficient for WPANs functionalizing at huge-data-rates. Therefore, a multi-hop provisional WPAN machine is appropriated at this juncture so as to maximize the coverage of UWB radio. Performance of the entire machine is achieved to determine if the Quality-of-Service conditions can, now even, be sustained when an IEEE 802.15.3 TDMA MAC stratum is used in multi-hop correspondence situations. Simulation outputs for Multi Hop WPAN standing on the UWB - OFDM physical layer are reproduced in this paper. In this mode of functioning, the transmitting machines for the data rates of 200 Mbps, 480 Mbps are used because these two are the directives for the highest compulsion rate and the greatest optional rate respectively. We used both 9mX 9m and 20mX20m geographical areas for the networks fields for the Multi Hop scenarios in this simulation model. The critical functionalities of the Multi Hop WPANS like average End – to – End Delay and Packet Failure Rate(PFR) and for all the source – Destination pairs are manipulated and restricted by employing the Qualnet network simulator.

Keywords- Multi hop, OFDM, Throughput, UWB, WPAN's

I. INTRODUCTION

At this juncture, there is a huge requirement for wireless communication systems that could be monitored at high amount of data rates over a very less distance communications so as to attain the modern advances in electronic gadgets (Camcorders, DVD Players, etc). The usage of high - rate Wireless Personal Area Networks (WPANs) for short distances provisional connectivity among electronic gadgets and communication devices have paved their way since 2000. having been approved from Federal Communications Commission (FCC) for the use of Ultra- Wide- Band (UWB) on the unlicensed band in 3.1 – 10.6 GHz range maximizes the

extensive usage of cutting edge WPAN networks (up to 480 Mbps) grounding on a UWB physical layer application. The special interest group (SIG) from IEEE have structured for this high- rate WPANS, which is popularly known as IEEE 802.15.3.

We begin with the thought of Multi Hop Wireless Personal Area Network (WPAN) in this paper, then the confrontations of the Multi Hop WPANS, and later the reflections of Multi Hop WPANS for the performance assessments like End- to- End delay, Packet Failure rate calculations for both the data rates of 200 Mbps and 480 Mbps.

II. MULTI HOP WPAN'S

Mobile multi-hop Adhoc networks (MANETs) are assortments of mobile nodes of bridges linked together over a wireless viaduct. These nodes can freely and actively self-monitor into approximate and temporary expedient network analysis sites. In this way, instruments can seamlessly inter-network in areas where pre-existing communication infrastructure (e.g., disaster recovery sites and battlefield environments) is zero. The discreet connectivity concept is not a budding one , but has been in existence for the last 30 years in different modes such as packet radio network (1972), sustainable adaptive radio network (1980), Global Mobile information system (early 1990s). Due to their quick and economically less demanding deployment of Ad hoc wireless networks we observe applications for the same in many areas. Defense applications, associated and spearheaded computing, emergency operations, wireless mesh networks, wireless sensor networks, and hybrid wireless network architectures are some of the areas its applications. Conventionally, logical networks have been the only correspondence networking practice that accepted the ad hoc paradigm. The thumb-rule behind provisional networking is that of multi-hop relaying.

In cellular networks, the routing decisions are acceded in a centralized format under the surveillance of base stations. But in an ad hoc cordless network, both accessing and resource management are operated in a scattered form in which all nodes would associate to capacitate communication among the nodes themselves. This calls for each bridge to be more

intelligible so that it can act both as a data signaling host for transmitting and receiving data, and as a network lane for routing packets from other ends. Hence, the mobile paths in possible wireless networks are more confusing and entangled than that of their correspondents in cellular networks. The truancy of any central administrator, or control station, makes the routing process a more complicated one compared to that found in cellular networks. Multi-network "hops" may be required for one station to interchange information with another node located elsewhere in the network due to the restricted transmission range of a wireless network. In such a network, each mobile node operates not only as a host but also as a router, forwarding packets for other mobile nodes in the network that may not be within direct wireless transmission range of each other. Each node involves in an accessing protocol that permits it to search for "Multi-hop" paths through the network to any other node.

WPAN is said to be a single-hop network as per the present IEEE 802.15.3 Strategy. That is, an info packet can be forwarded only from a source address to a destination address, and there is no arbitrating node to work as a "router". Using an UWB - OFDM physical layer practicability for a WPAN, the amount that can be attained is acutely minute, usually less than 10 meters. For an assured transmission with minimal packet error progression, a certain concentration of within 4 meters is usually needed. The benefit with a multi-hop network is obvious as it can maximise network coverage without increasing either the accessibility strength, or sensitivity of the receiver. The other advantage is that of improved reliability through redundancy of route. The ambit of IEEE 802.15.3 MAC code to provide multi-hop networks calls for attentive and comprehensive observation.

An example is used to demonstrate why a Multi - hop WPAN is required to provide backup for immense progression practical traffic flows. A video conference or home theatre system is a trivial practice for use of WPAN based on the OFDM UWB physical layer. That is, to transmit the multimedia traffic instead of using cables, the unwired links will be used. The frequency range requirements for each traffic outflow is about 6 Mbps, the average downtime should be less than 90 ms, and the packet Failure rate, less than 8% so as to arrive at the required QoS level. The circuitry region for a video conference or home theatre system generally ranges from 9 m x 9 m to 20 m x 20 m. The indemnity radius for an UWB - OFDM regulation is relatively only 3 meters for a data procession of 200 Mbps and only 7 meters for a info progression of 480 Mbps to guarantee a PER of 8%. A single-hop network structure is inadequate to cover the expected network area for these huge amounts of data rates have retained obvious. If a Multi - hop WPAN frame works well, then the network coverage area can be perfectly enlarged through the application of arbitrary nodes while monitoring transmission at the required data rates. The suitability of the IEEE 802.15.3 TDMA MAC layer for use with multi-hop WPAN systems necessitates to be recognized. In Multi - hop

network, due to the huge amount of variables taken part, the amplitude of the machine develops significantly, thus materializing logical modeling a considerably arduous task. On the side of the machine, simulation methods capacitate the exploration of more problematic and realistic phenomena. In composite machinery such as multi-hop networks, attentive preference of the system attributes can drive to considerable development in function, specifically for time-sensitive applications. Focusing on time-sensitive applications, the objective is to examine the performance strategies of multi-hop WPAN systems standing on an OFDM physical layer. Compatible system functioning precautions involving end-to-end delay, productivity and packet failure rate realized in various conditions with different choices of system parameters.

A. Capacity Analysis of a Multi-Hop Network

The network productivity or approximate capacity for a multi-hop network is described in this section. When frequency reuse is not considered, the capacity of multi-hop networks is greatly affected by the average hop count h . Theoretically, if the network capacity based on peer-to-peer communications is C , the capacity of multi-hop networks will be $C = C/h$, assuming that the network bandwidth used for routing messages is multi negligible, and that a high-efficiency scheduling scheme is implemented. If the aggregate packet production rate is r Mbps, the highest number of source-destination pairs that can be supported is $L = C/r$. When the number of source-targeted pairs L is max multi over L , packets will be launched due to the existence of a network due point condition at max.

The conversion and transformation system being monitored at 200 Mbps is utilized here to exemplify how the Multi- hop network ability is related to the associated network strategy and the average hop count. It is known that the attainable productivity for 200 Mbps peer-to-peer transmission is about 120 Mbps. If the average hop count is set to $h = 3$, the capacity of a multi-hop network will be $C = 120/3 = 40$ Mbps, theoretically. the maximum number of source-destination = $40/6 = 6$, if the average packet multi generation rate per link is $r = 6$ Mbps,. If the packet Generation rate doubles, that can be supported is $L = C/r$ max multi per link $r = 3$ Mbps, then the maximum number of source-destination pairs that can be = $40/3 = 13$. If the average hop count is fixed to backed up is $L = C/r$ $h = 4$, the max multi capacity of a multi-hop network will be $C = 120/4 = 30$ Mbps, theoretically. If the multi average packet generation rate per link is $r = 6$ Mbps, then the maximum number of source-destination pairs that can be supported is $L = C/r = 30/6 = 5$. The maximum number of source-destination pairs that can be backed up is $L = C/r = 30/3 = 10$. If the max multi average packet generation rate per link is $r = 3$ Mbps. When the max multi number of source-estimation pairs L is greater than L , packets will get a break down affected due to the saturation of max network.

Resultantly, the packet failure rate and the aggregate downtime should increase productively.

III. PREVAILING CHALLENGES IN MULTI-HOP NETWORKS

In a multi-hop provisional network, connections correspond with each other using multi-hop wireless links, and there are no static infrastructure instruments similar to a ground station. Each connection in the network also plays a role as a router, enrouting data packets for other nodes. One of the prominent hurdles is the structure of active routing protocols that can efficiently search for routes between two corresponding nodes. Routing is apparently the first methodology to be reconsidered in altering from single-hop to multi-hop implementations [6]. A mobile ad hoc networking (MANET) functioning set has been established within the Internet Engineering Task Force (IETF) to develop a routing framework for IP-based protocols in ad hoc networks. Dozens of routing protocols for MANETs have been introduced, some examples including DSDV (Destination Sequenced Distance Vector), DSR (Dynamic Source Routing), and AODV (Ad-hoc On-demand Distance Vector). However, most simulations and performance affinities of mobile Adhoc network piloting protocols are based on a condensed and visionary physical layer model, as well as easy performance metrics.

Most of the presently prevailing codes were framed out under the hypothesis of an UDG (Unit Disk Graph) communication model, in which signal strength variations due to a realistic channel are not considered. Without modification, such routing schemes cannot work well with physical layer characteristics that are correspondent of more factual communication channel environments.

IV. SIMULATION RESULTS FOR MULTI-HOP WPAN SYSTEMS

The simulation results for multi-hop communication system structuralizing are exhibited, and the assistive performance analyses are given in this paper. The transmission systems operating at 200 Mbps and 480 Mbps are simulated in this analysis as they are representatives of the immense mandatory rate and the immense optional rate, respectively. First, the simulation results and function analysis for the equal-weighted node-based scheduling scheme are shown. Then, the simulation outputs and performance analysis for the on-demand link-based scheduling scheme are given.

In an unorthodox simulation scheme we applied for Multi – Hop networks are basically depended on the Link formation algorithm because of the existence of direct relationship between the Throughput and the scheduling competence. In this imaging task we used the two Link organizing algorithms; the first is Equal-Weighted Node-Based Scheduling and the second, On-Demand Link-Based Scheduling.

Table 1. abridges the limitations of system used in the simulations for the Multi -hop situations recognized in this analysis.

Simulation parameter	Value
Simulation Time	5s
Number of nodes	20
Number of links	2,4,6,8,10
Network Area	20mX20m for 200 Mbs 9mX9m for 480 Mbs
Node's coverage radius to achieve a PER of 5%	6.9m for 200 Mbs 2.95m for 480 Mbs
Number of Channels	1(Center Frequency = 3.432 GHz
Transmission Power	-10.3 dBm
Receiver sensitivity	-77.2 dBm for 200 Mbs -72.6 dBm for 480 dBm
Channel model considered	Free space,Shadowing,and Rayleigh fading
Packet size(application layer)	982 bytes(will be 1024 bytes after MAC layer)
Max Network Buffer size	1,00,000 Bytes
CTA slot Duration	Transmission duration of 1024-Byte Packet
Number of slots per Frame for Equal- Weighted Node-Based Scheduling	20
Number of slots per Frame for On – Demand Link-Based Scheduling	20,40 for 200 Mbs 30,60 for 480 Mbs
Guard time between slots	1 μ s
Intra Frame time	1.875 s

A. SIMULATION RESULTS FOR EQUAL-WEIGHTED NODE-BASED SCHEDULING

The equal-weighted node-based scheduling scheme is first implemented. The packet generation rates are taken to be 128 kbps, 3 Mbps and 6 Mbps. Figures 1 and 2 exemplify the average delay and the PFR with PGR taken as a parameter using the equal- weighted scheduling scheme for systems operating at 200 Mbps. Figures 3 and 4 illustrate the average delay and the PFR with PGR considered a parameter using the equal- weighted scheduling scheme for systems being operated at 480 Mbps.

Each node has the same share of the bandwidth irrespective of whether it has a packet to transmit or not and independent of how many packets it needs to transmit for equal-weighted node-based scheduling. For the total number of network nodes set to 20, each node can have $120/20 = 6$ Mbps of frequency of the network available for systems being operated at 200 Mbps, and $180/20 = 9$ Mbps of network bandwidth available for systems operating at 480 Mbps. If the PGR per link is 6 Mbps, only 1, or possibly 1.5 traffic currents can be backed up by one node in either case. So, there will be collisions, and some of the packets will be dropped, if a node is a transmitting node for one traffic progression and a forwarding node for another traffic stream. This situation occurs rarely, and sometimes there are number of traffic currents which need to be

transmitted by one node at the same time. Hence, the system may work well with high probability only when the number of source-destination pairs is very small. When there are not more than 2 active links when the PGR equals 6 Mbps for systems operating at either 200 Mbps, or 480 Mbps only, the simulation results show that the performance measures are acceptable. When the number of source-destination pairs L is greater than 2, both the PFR and the average delay increase logically. Similarly, if the PGR per link is 3 Mbps, only 2 or 3 traffic streams can be transmitted from one node at the same time in either case. The situation is better than that for a PGR equal to 6 Mbps, but the capacity available for each node is still not enough. It can be observed that a maximum of 4 active links can be supported. When $L > 4$, both the PFR and the delay maximizes dramatically. The maximum numbers of source-destination pairs that can be supported are less than the theoretically predicted capacities that were presented in Section II.A for machines being operated at either 200 Mbps, or 480 Mbps. The efficiency of allotment is less, and the system bandwidth is wasted. For a PGR equal to 128 kbps, there are over 50 traffic currents that can be backed by any one node at the same time for systems operating at either 200 Mbps, or 480 Mbps. when the PGR is 128 kbps, it can be recorded that the PFR (<8%) and the delay (about 5ms) both meet the QoS requirements for real-time applications even for 10 active links. The Equal - weighted scheduling scheme only works well when either the packet generation rate is low, or there is only a very small number of active links. However, a UWB-based WPAN system is structured for high-data rate inter media progression, and hence, QoS requirements have to be met. The simple equal- weighted node-based scheduling cannot execute well in this kind of condition. For huge amount of info speeds, the on-demand scheduling scheme has to be considered.

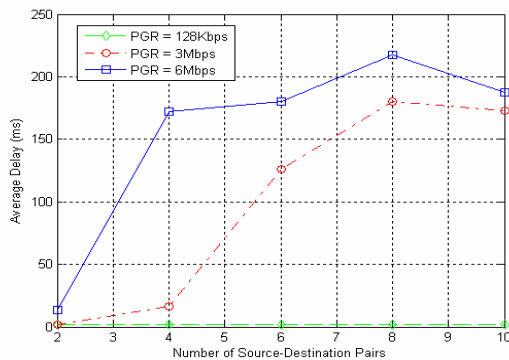


Figure 1.: Average Delay vs. Number of Source-Destination Pairs With Equal- Weighted Scheduling for Transmission Systems Operating at 200 Mbps

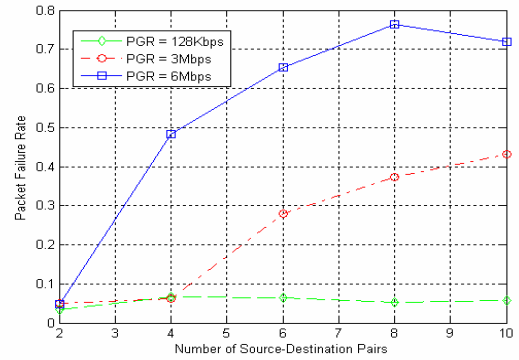


Figure 2.: PFR vs. Number of Source-Destination Pairs With Equal-Weighted Scheduling for Transmission Systems Operating at 200 Mbps.



Figure 3.: Average Delay vs. Number of Source-Destination Pairs With Equal-Weighted Scheduling for Transmission Systems Operating at 480 Mbps

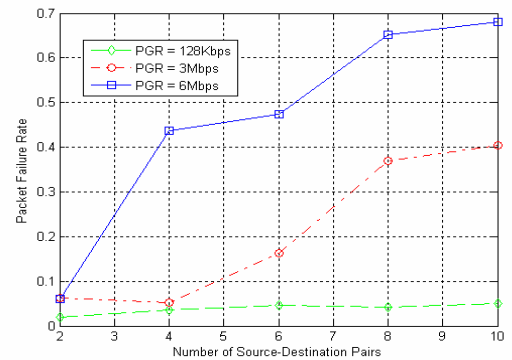


Figure 4.: PFR vs. Number of Source-Destination Pairs With Equal-Weighted Scheduling for Transmission Systems Operating at 480 Mbps

B. SIMULATION RESULTS FOR ON-DEMAND LINK-BASED SCHEDULING

For the on-demand link-based scheduling scheme, the packet generation rates are absorbed to be 3 Mbps and 6 Mbps. A value for PGR of 128 Kbps is not accepted here for the on-demand link- based scheduling scheme, provided that the equal-weighted scheduling can function perfectly for low data rates.

As the criteria of using the on-demand link-based scheduling scheme for systems operating at 200 Mbps, Figures 5 and 6 explain the aggregate delay and the PFR with PGR, respectively. It can be marked that saturation of the network is reached when there are more than 6 dynamic connections for a PGR similar to 6 Mbps. Both the PFR ($< 7\%$) and the delay (< 40 ms) are appropriated for real-time applications before network due-point happens. Another analysis is that both the PFR ($< 7\%$) and the delay (< 40 ms) are feasible even for the case of 10 dynamic links when the PGR is 3 Mbps per link.

These simulation yields for systems operating at 200 Mbps match the theoretically assumed capacities that were shown in Section II.A. That is, a total of 6 links can be reinforced when the PGR is equal to 6 Mbps and 12 links can be supported when the PGR is equal to 3 Mbps. Figures 7 and 8 exemplify the average delay and the PFR, respectively, using the needed scheduling scheme for systems being functioned at 480 Mbps. It can be considered that saturation of the network is attained when there are more than 8 active links for a PGR equal to 6 Mbps. Both the PFR ($< 7\%$) and the delay (< 10 ms) remain reasonable before network saturation occurs.

Another observation is that both the PFR ($< 7\%$) and the delay (< 10 ms) are acceptable even for the case of 10 active links when the PGR is 3 Mbps per link. The simulation results attained for networks functionalizing at 480 Mbps match the theoretically and impractically assumed capacities that were produced in Section II.A. That is, 8 links can be upheld when the PGR is equal to 6 Mbps and 16 links can be supported when the PGR is equal to 3 Mbps.

When the PGR is 3 Mbps per link, this will also be examined that both the PFR and the delay reach the QoS requirements for real-time applications even for 10 active links. With the same network buffer size, the PFR is almost the same when the PGR is equal to 6 Mbps and when the PGR is equal to 3 Mbps. The delay when the PGR is same as to 3 Mbps which is slightly smaller than that when the PGR is equal to 6 Mbps. This is feasible since there will be more adjoining deferment associated with the higher data rate.

The simulation outputs described above for machines being monitored at both 200 Mbps and 480 Mbps match the capacity analysis for a multi-hop network exhibited in Section II.A. Hence, it can be examined that the efficiency in allotment is comparatively greater for the required scheduling scheme, and the network bandwidth can be utilized more efficiently than in the case of the equal-weighted scheduling scheme. It can be summarized that this UWB-based multi-hop WPAN system performs well when the on-demand link-based scheduling is used along with the proper routing protocol.

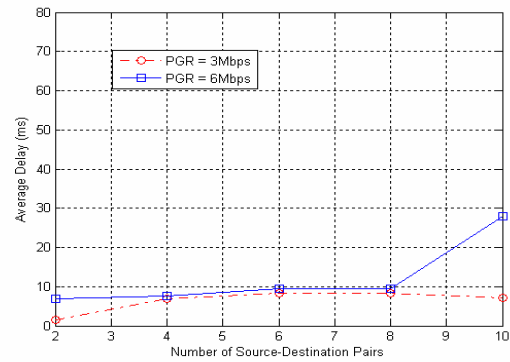


Figure 5.: Average Delay vs. Number of Source-Destination Pairs With On-Demand Scheduling for Transmission Systems Operating at 200 Mbps

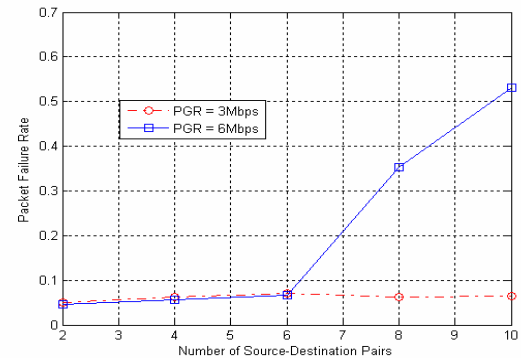


Figure 6.: PFR vs. Number of Source-Destination Pairs With On-Demand Scheduling for Transmission Systems Operating at 200 Mbps

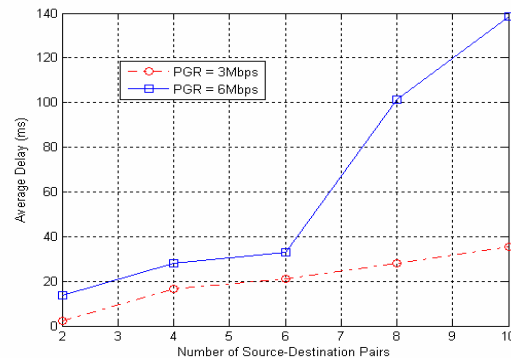


Figure 7.: Average Delay vs. Number of Source-Destination Pairs With On-Demand Scheduling for Transmission Systems Operating at 480 Mbps

III CONCLUSIONS

Based on the simulation results attained and performance analyses described in the previous section, conclusions can be drawn. The equal-weighted node-based allotting scheme does not function well for high-data rate applications. That is, the

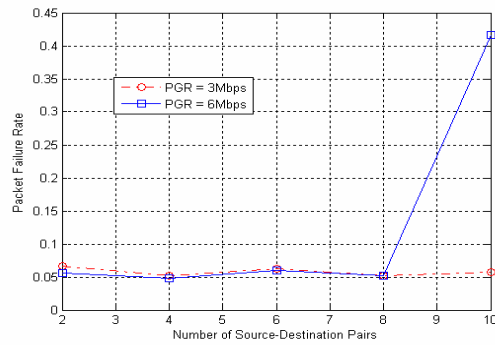


Figure 8.: PFR vs. Number of Source-Destination Pairs With On-Demand Scheduling for Transmission Systems Operating at 480 Mbps.

scheduling efficiency is low and much of the available network frequency had been wasted. When either the data rate is very low, or there are only a very small number of active links, this scheduling scheme only executes well since the network bandwidth is not utilized efficiently.

The On - Demand link-based scheduling scheme can perform well for the UWB-based multi-hop WPAN system taken into view here. That is, the scheduling efficiency is high, and the network bandwidth is utilized efficiently. Thus, the IEEE 802.15.3 TDMA MAC layer with the accurate scheduling and routing schemes perform well in the context of multi-hop networks. Multi-hop WPANs based on a realistic OFDM UWB physical layer can be a suitable method to improve the network coverage while backing up huge amount of data rate multimedia traffic.

REFERENCES

- [1] R. Bruno, M. Conti and E. Gregori, "Mesh Networks: Commodity Multihop Ad Hoc Networks," IEEE Communications Magazine, pp. 123-131, March 2005.
- [2] C. S. Murthy and B. S. Manoj, Ad Hoc Wireless Networks: Architecture and Protocols, Prentice-Hall, NJ, 2004.
- [3] F. Eshghi, A. K. Elhakeem and Y. R. Shayan, "Performance Evaluation of Multihop Ad Hoc WLANs," IEEE Communications Magazine, pp. 107-115, March 2005.
- [4] A. F. Molisch, J. R. Foerster and M. Pendergrass, "Channel Models for Ultrawideband Personal Area Network," IEEE Wireless Communications Magazine, Vol. 10, pp. 14-21, December 2003.
- [5] M. D. Benedetto and G. Giancola, Understanding Ultra Wide Band Radio Fundamentals, Prentice-Hall, NJ, 2004.
- [6] A. Saleh and R. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," IEEE Journal on Selected Areas in Communications, Vol. 11, No. 7, pp. 967-978, September 1993.
- [7] L. Maret, I. Siaud and Y. Kamiya, "Ultra WideBand PHY Layer MBOA Performance and Sensitivity to Multipath Channels (IST Magnet Project)," <http://www.ist-magnet.org/>.
- [8] MultiBand OFDM Alliance, "Multi-Band OFDM Physical Layer Proposal for IEEE 802.15 Task Group 3a," September 14, 2004, <http://www.wimedia.org/>.
- [9] H. Xu and A. Ganz, "A Radio Resource Control Method in UWB Protocol Design," Military Communications Conference, Vol. 2, pp. 886-891, October 2003.
- [10] S. Datta, I. Seskar and M. Demirhan, "Ad-hoc Extensions to the 802.15.3 MAC Protocol," Proceedings of the Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM'05), Taormina, Giardini Naxos, pp. 293-298, June 2005.
- [11] A. Rangnekar and K. Sivalingam, "Multiple Channel Scheduling in UWB Based IEEE 802.15.3 Networks," Proceedings of the First International Conference on Broadband Networks (BROADNETs), San Jose, CA, pp. 406-415, October 2004.
- [12] H. Fattah and C. Leung, "An Overview of Scheduling Algorithms in Wireless Multimedia Networks," IEEE Wireless Communications Magazine, pp. 76-83, October 2002.
- [13] I. Stojmenovic, A. Nayak and J. Kuruvila, "Design Guidelines for Routing Protocols in Ad Hoc and Sensor Networks with a Realistic Physical Layer," IEEE Communications Magazine, pp. 101-106, March 2005.
- [14] H. Gao and D. G. Daut, "Position-Based Greedy Stateless Routing for Multihop WPANs Based on a Realistic UWB Physical Layer," Second IEEE International Conference on Wireless Communications, Networking, and Mobile Computing (WiCOM), Wuhan, P. R. China, September 2006.
- [15] D. Couto, D. Aguayo, J. Bricket and R. Morris, "A High-Throughput Path Metric for Multi-Hop Wireless Routing," International Conference on Mobile Computing and Networking (MobiCom), San Diego, CA, pp. 134-146, September 2003.
- [16] H. Tsai, N. Wisitpongphan and O. K. Tonguz, "Link-Quality Aware Ad Hoc On- Demand Distance Vector Routing Protocol," First International Symposium on Wireless Pervasive Computing, Phuket, Thailand, January 2006.
- [17] L. Qin and T. Kunz, "On-demand Routing in MANETs: The Impact of a Realistic Physical Layer Model," Proc. Second International Conference on Ad Hoc, Mobile and Wireless Networks, Montreal, Canada, pp. 37-48, October 2003.
- [18] S. Lee, B. Bhattacharjee and S. Banerjee, "Efficient Geographic Routing in Multihop Wireless Networks," International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Urbana-Champaign, IL, pp. 230-241, May 2005.

AUTHORS PROFILE



Prof. Ch. Subrahmanyam presently working as a Professor & Head, Department of ECE at Scientist Institute of technology, Hyderabad. He has completed his B.E. in 1995 from Andhra University, A. P. India, and M. Tech. from JNTU Hyderabad, in 2002, and Pursuing his Ph.D. from JNTU Hyderabad, A. P. India with ECE in Wireless communications. He has about 15 years of experience in teaching and industry together, he is having publications in International Journals and Conferences. He has guided many M. Tech and B. Tech. Projects. He is a life member of ISTE, India.



Dr. K. Chennakeshava Reddy, Presently working as Principal & Professor of ECE at TKR College of Engineering. He has completed his B.E. in 1973 and M. Tech. in 1976 from REC Warangal, A.P. India, and Ph.D. in 2001 from JNTU Hyderabad. He has worked in various positions starting from lecturer to Director of Evaluation in JNT University, Hyderabad, A. P. India. He has about 70 publications in international and National journals and Conferences and he has successfully guided 4 Ph.Ds and many are under progress. He is a member of various technical Associations.



Dr. Syed Abdul Sattar, presently working as a Dean of Academics & Professor of ECE department, RITS, Chevella, Hyderabad. He has completed his B.E. in ECE in 1990 from Marathwada university Aurangabad, M.S. India, M. Tech. In DSCE from JNTU Hyderabad, in 2002, and Pursued his first Ph.D. from Golden state University USA, with Computer Science in 2004, and second Ph.D. from JNTU Hyderabad, A. P. India with ECE in 2007. His area of specialization is wireless communications and image Processing. He has about 21 years of experience in teaching and industry together and recipient of national award as an Engineering Scientist of the year 2006 by NESA New Delhi, India. He has about 73 publications in International and National Journals and conferences. Presently he is guiding more than 15 research scholars in ECE and Computer Science from different Universities. He is a member of Board of studies for a central university and reviewer/editorial member/chief editor for national and International journals.

Face Recognition Using Biogeography Based Optimization

Er. Navdeep Kaur Johal
R.I.E.I.T.
Railmajra, Distt. SBS Nagar
Punjab, India
navdeepkjohal@gmail.com

Er. Poonam Gupta
R.I.E.I.T.
Railmajra, Distt. SBS Nagar
Punjab, India
poonamjindal3@gmail.com

Er. Amanpreet Kaur
R.I.E.I.T.
Railmajra, Distt. SBS Nagar
Punjab, India
amanrandhawa85@gmail.com

Abstract: Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. It is the most important step that affects the performance of a pattern recognition system. This paper presents a novel feature selection algorithm based on Biogeography Based Optimization (BBO). Biogeography-based optimization (BBO) is a recently-developed EA motivated by biogeography, which is the study of the distribution of species over time and space. The algorithm is applied to coefficients extracted by discrete cosine transforms (DCT). The proposed BBO-based feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully selected according to a well defined discrimination criterion. Evolution is driven by a fitness function defined in terms of maximizing the class separation (scatter index). The classifier performance and the length of selected feature vector are considered for performance evaluation using the ORL face database. Experimental results show that the BBO-based feature selection algorithm was found to generate excellent recognition results with the minimal set of selected features.

Keywords: Face Recognition, Biogeography Based Optimization, DCT, Feature Selection

I. INTRODUCTION

Face Recognition is a process in which we match the input image with the given database and produce the output image which is similar to the input image. As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past several years. At least two reasons account for this trend: the first is the wide range of commercial and law enforcement applications, and the second is the availability of feasible technologies after 30 years of research. Even though current machine recognition systems have reached a certain level of maturity, current systems are still far away from the capability of the human perception system. So many

approaches to face recognitions have been developed; an excellent survey paper on the different face recognition techniques can be found in [1].

A. Feature Extraction

The first step in any face recognition system is the extraction of the feature matrix. A typical feature extraction algorithm tends to build a computational model through some linear or Non-linear transform of the data so that the extracted feature is as representative as possible or when the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called *feature extraction*. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Best results are achieved when an expert constructs a set of application-dependent features. Nevertheless, if no such expert knowledge is available general dimensionality reduction techniques or feature extraction may help. These include:

- geometrical features extraction
- statistical (algebraic) features extraction [2 - 8].

The geometrical approach, represent the face in terms of structural measurements and distinctive facial features that include distances and angles between the most characteristic face components such as eyes, nose, mouth or facial templates such as nose length and width, mouth position, and chin type. These features are used to recognize an unknown face by matching it to the nearest neighbor in the stored database. Statistical features extraction is usually driven by algebraic

methods such as principal component analysis (PCA), and independent component analysis (ICA) [6]. These methods find a mapping between the original feature spaces to a lower dimensional feature space.

Alternative algebraic methods are based on transforms such as downsampling, Fourier transform (FT), discrete cosine transform (DCT), and the discrete wavelet transform (DWT). Transformation based feature extraction methods such as the DCT was found to generate good FR accuracies with very low computational cost [8].

B. Discrete Cosine Transform

DCT has emerged as a popular transformation technique widely used in signal and image processing. This is due to its strong “energy compaction” property: most of the signal information tends to be concentrated in a few low-frequency components of the DCT. The use of DCT for feature extraction in FR has been described by several research groups [9-15]. DCT transforms the input into a linear combination of weighted basis functions. These basis functions are the frequency components of the input data.

The general equation for the DCT of an NxM image $f(x, y)$ is defined by the following equation:

$$F(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \cos\left[\frac{\pi u}{2N}(2x+1)\right] \cos\left[\frac{\pi v}{2M}(2y+1)\right] f(x, y) \quad \dots (i)$$

Where $f(x, y)$ is the intensity of the pixel in row x and column y ; $u = 0, 1, \dots, N-1$ and $v = 0, 1, \dots, M-1$ and the functions $\alpha(u)$, $\alpha(v)$ are defined as:

$$\alpha(u), \alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } u, v = 0 \\ \sqrt{\frac{2}{N}} & \text{for } u, v \neq 0 \end{cases} \quad \dots (ii)$$

For most images, much of the signal energy lies at low frequencies (corresponding to large DCT coefficient magnitudes); these are relocated to the upper-left corner of the DCT array. Conversely, the lower-right values of the DCT array represent higher frequencies, and turn out to be small enough to be truncated or removed with little visible distortion. This means that the DCT is an effective tool that can pack the most effective features of the input image into the fewest coefficients.

C. Feature Selection

After extracting the features, we further need minimal subset of features so that we are able to recognize the face. Due to this reason we need a feature selection algorithm that reduces the

maximum number of irrelevant and redundant features obtained during feature extraction while maintaining acceptable classification accuracy. Among the various methods proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method [16-18] and Ant Colony Optimization (ACO)-based method have attracted a lot of attention [19]. In the proposed FR system we utilized an evolutionary feature selection algorithm based on swarm intelligence called the Biogeography Based Optimization. Biogeography Based Optimization is explained in the next section.

D. Biogeography based Optimization

Biogeography is the study of the distribution of biodiversity over space and time. It aims to analyze where organisms live, and in what abundance. Biogeography is modeled in terms of such factors as habitat area and immigration rate and emigration rate, and describes the evolution, extinction and migration of species. Biogeography-Based Optimization (BBO) is a new biogeography inspired algorithm for global optimization. BBO [20] is a new biogeography inspired global optimization algorithm, which is similar to the island model-based GAs [21]. Each individual is considered as a “habitat” with a habitat suitability index (HSI) to measure the individual. The variables of the individual that characterize habitability are called suitability index variables (SIVs). In BBO, each individual has its own immigration rate λ and emigration rate μ . The immigration rate and emigration rate are functions of the number of species in the habitat. They can be calculated as follows:

$$\lambda_k = I \left(1 - \frac{k}{n} \right) \quad \dots (iii)$$

$$\mu_k = E \left(\frac{k}{n} \right) \quad \dots (iv)$$

where I is the maximum possible immigration rate; E is the maximum possible emigration rate; k is the number of species of the k th individual; and n is the maximum number of species. Note that Eqs. (iii) and (iv) are just one method for calculating λ and μ , there are other different options to assign them based on different species models [20].

In BBO, there are two main operators, i.e., migration and mutation. Suppose that we have a global optimization problem and a population of candidate individuals. The individual is represented by a D-dimensional integer vector (SIV). The population consists of $NP = n$ parameter vectors X_i , $i = 1 \dots NP$.

One option for implementing the migration operator and the mutation operator can be described in Figure 1 and 2, respectively. Where $\text{rndreal}(0, 1)$ is a uniformly distributed random real number in $(0,1)$ and $X_i(j)$ is the j th SIV of the solution X_i . m_i is the mutation rate that is calculated as:

$$m_i = m_{\max} \left(1 - \frac{P_i}{P_{\max}} \right) \quad \dots(v)$$

where m_{\max} is an user-defined parameter, and $P_{\max} = \arg \max P_i$, $i = 1, \dots, NP$. Each population member has an associated probability, which indicates the likelihood that it was expected *a priori* to exist as a solution to the given problem. The steady state value for the probability of the number of each species to exist is given by [22]:

$$P_k = \begin{cases} P_0 = \frac{1}{1 + \sum_{k=1}^n \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k}}, & k=0 \\ P_k = \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k \left(1 + \sum_{k=1}^n \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k} \right)}, & 1 \leq k \leq n \end{cases} \quad \dots(vi)$$

The largest possible number of species that the habitat can support is n . It is necessary that $\mu_k \neq 0$ for all k for this limiting probabilities to exist.

```

1: for i = 1 to NP
2:   Select  $X_i$  with probability  $\propto \lambda_i$ 
3:   if  $\text{rndreal}(0, 1) < \lambda_i$  then
4:     for j = 1 to NP do
5:       Select  $X_j$  with probability  $\propto \mu_j$ 
6:       if  $\text{rndreal}(0, 1) < \mu_j$  then
7:         Randomly select an SIV  $\sigma$  from  $X_j$ 
8:         Replace a random SIV in  $X_i$  with  $\sigma$ 
9:       end if
10:    end for
11:  end if
12: end for

```

Figure.1: Algorithm for Habitat Migration.

```

1: for i = 1 to NP
2:   Compute the probability  $P_i$ 
3:   Select SIV  $X_i(j)$  with probability  $\propto P_i$ 
4:   if  $\text{rndreal}(0, 1) < m_i$  then
5:     Replace  $X_i(j)$  with a randomly generated SIV
6:   end if
7: end for

```

Figure.2: Algorithm for Habitat Mutation

With the migration operator, BBO can share the information between solutions. A good solution is analogous to an habitat

with a high HSI, and a poor solution represents an habitat with a low HSI. High HSI solutions resist change more than low HSI solutions. By the same token, high HSI solutions tend to share their features with low HSI solutions. (This does not mean that the features disappear from the high HSI solution; the shared features remain in the high HSI solutions, while at the same time appearing as new features in the low HSI solutions. This is similar to representatives of a species migrating to a habitat, while other representatives remain in their original habitat). Poor solutions accept a lot of new features from good solutions. This addition of new features to low HSI solutions may raise the quality of those solutions. Good solutions have high emigration rate and they share their features (SIVs) with bad solutions that have high immigration rate. Additionally, the mutation operator tends to increase the diversity of the population. The BBO algorithm can be described with the following algorithm in figure 3:

Pseudo-code for biogeography-based optimization. Here H indicates habitat, HSI is fitness, SIV (suitability index variable) is a solution feature, λ denotes immigration rate and μ denotes emigration rate.

Biogeography-Based Optimization (BBO)

Begin

/* BBO parameter initialization */

1. Create a random set of habitats (population)

H_1, H_2, \dots, H_n ;

2. Compute corresponding HSI values;

/* End of BBO parameter initialization */

3. While not T /* T is a termination criterion */

4. Compute immigration rate λ and emigration rate μ for each habitat based on HSI;

/* Migration */

5. Apply migration as defined in algorithm 1.

/* End of migration */

/* Mutation */

6. Apply mutation as defined in algorithm 2.

/* End of mutation */

7. Recompute HSI values;

8. End while

9. End

Figure 3: Main BBO Algorithm

II. BBO-BASED FEATURE SELECTION

In this proposed work, features of image are extracted using DCT technique. The extracted features are reduced further by using Biogeography Based Optimization to remove redundancy and irrelevant features. The resulting feature subset (obtained by BBO) is the most representative subset and is used to recognize

the face from face gallery.

A. Habitat Representation

In proposed work, each habitat represents one possible solution (feature subset) required for face recognition. Each of the features extracted by DCT of image represents one Suitability Index Variable (SIV) of the habitat. Further, during feature subset selection each of these feature is either selected or rejected, SIV C is an integer and $C \in [0, 1]$. A habitat $H \in SIV^m$ where m is the length of the feature vector extracted by the DCT.

B. SIV Mutation

In proposed work, a habitat is chosen for mutation based on mutation rate and species count probabilities defined in (4) and (5). Once a habitat is selected for mutation, a random SIV is selected; it is mutated to 0 if its value is 1 or vice versa. Therefore, if a particular feature was earlier selected, it is rejected after mutation and vice versa.

C. Habitat Suitability Index

In each generation, each habitat is evaluated, and a value of *goodness* or *fitness* is returned by a fitness function. This evolution is driven by the fitness function F that evaluates the quality of habitat in terms of their ability to maximize the class separation term indicated by the scatter index among the different classes [23]. Let w_1, w_2, \dots, w_L and N_1, N_2, \dots, N_L denote the classes and number of images within each class, respectively. Let M_1, M_2, \dots, M_L and M_0 be the means of corresponding classes and the grand mean in the feature space, M_i can be calculated as:

$$M_i = \frac{1}{N_i} \sum_{j=1}^{N_i} W_j^{(i)}, \quad i = 1, 2, \dots, L \quad \dots \text{(vii)}$$

Where $W_j^{(i)}$, $j=1, 2, \dots, N_i$, represents the sample image from class w_i and grand mean M_0 is:

$$M_0 = \frac{1}{N} \sum_{i=1}^L N_i M_i \quad \dots \text{(viii)}$$

Where N is the total no of images of all the classes. Thus the between class scatter fitness function F is computed as follow:

$$F = \sqrt{\sum_{i=1}^L (M_i - M_0)' (M_i - M_0)} \quad \dots \text{(ix)}$$

D. Classifier

After the training phase, a typical and popular Euclidean distance is employed to measure the similarity between the test vector and the reference vectors in the gallery. Euclidean distance is defined as the straight-line distance between two points. For N -dimensional space, the Euclidean distance between two any points' p_i and q_i is given by:

$$D = \sqrt{\sum_{i=1}^N (p_i - q_i)^2} \quad \dots \text{(x)}$$

Where p_i (or q_i) is the coordinate of p (or q) in dimension i .

E. Proposed BBO-Based Feature Selection Algorithm

In the proposed work (figure 4), the features of image are extracted using DCT technique. These extracted features are further reduced (or selected) using BBO. In BBO, each SIV of habitats is randomly set to either 0 or 1 initially, which implies that initial feature subset selection is done randomly but after the completion of BBO algorithm, BBO helps to select the optimal set of features from the given features. The stopping criterion of proposed algorithm is number of iterations. At the end of training phase, we have the optimal set of features. These features are then selected from the test image and the face gallery. The test image is recognized as that face from face gallery which has minimum Euclidean distance from the test image on the basis of these selected features.

1. **Feature Extraction:** Obtain the DCT array by applying Discrete Cosine Transformation to image.
2. Take the most representative features of size $n \times n$ from upper left corner of DCT Array.
3. **Feature Selection:** Apply the BBO algorithm defined in algorithm 3 to obtain the feature subset of the extracted features.
4. Pick up the habitat H with *max* (HSI) value. The SIVs of this habitat H represent the best feature subset of the features defined in step 2.
(Feature Selection Ends)
5. **Classification:** calculate the difference between the feature subset (obtained in step 4) of each image of facial gallery and the test image with the help of Euclidean Distance defined in formula (x). The index of the image which has the smallest distance with the image under test is considered to be the required index.

Figure 4: Face Recognition using BBO based Feature Selection

III. EXPERIMENTAL RESULTS

The performance of the proposed feature selection algorithm is evaluated using the standard Cambridge ORL gray-scale face

database. The ORL database of faces contains a set of face images taken between April 1992 and April 1994 at the AT&T Laboratories (by the Oliver Research Laboratory in Cambridge, UK) [24] and [25]. The database is composed of 400 images corresponding to 40 distinct persons. The original size of each image is 92x112 pixels, with 256 grey levels per pixel. Each subject has 10 different images taken in various sessions varying the lighting, facial expressions (open/ closed eyes, smiling/ not smiling) and facial details (glasses/ no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). Four images per person were used in the training set and the remaining six images were used for testing.

TABLE I. BBO parameter setting

Size of ecosystem (No of Habitats)	30
Number of iterations of BBO algorithm	100
SIV value	0 or 1

In this work, we test the BBO-based feature selection algorithm with feature vectors based on various sizes of DCT coefficient. The 2-dimensional DCT is applied to the input image and only a subset of the DCT coefficients corresponding to the upper left corner of the DCT array is retained. Subset sizes of 50x50, 40x40, 30x30 and 20x20 of the original 92x112 DCT array are used in this work. Each of 2- dimensional subset DCT array is converted to a 1-dimensional array using raster scan. This is achieved by processing the image row by row concatenating the consecutive rows into a column vector. This column vector is the input to the subsequent BBO-feature selection algorithm.

To calculate average recognition rate for each problem instance (20X20, 30X30, 40X40, and 50X50 DCT Array), test image is randomly chosen from 40 classes. Five trials are taken for each problem instance. Average recognition is measured by knowing how many times correct faces were identified out of 5 trials (for each problem instance). The average recognition rate is measured together with the CPU training time and the average number of selected features for each problem instance. The algorithm has been implemented in Matlab 7 and the result for each problem instance is shown in table 2:

TABLE II. Results of BBO-FS algorithm

DCT Feature Vector Size	Number of Features input to BBO-FS	Average no. of features selected by BBO - FS	Training time (in seconds)	Average Recognition Rate
20X20	400	219	85.743	100%
30X30	900	451	95.166	100%
40X40	1600	814	136.178	100%
50X50	2500	1243	165.101	100%

For each of the problem instance (20X20, 30X30, 40X40, and 50X50), algorithm is run 5 times and each time, random test image is chosen to be matched with face gallery. The test face matches with image in face gallery in each trial and average recognition rate is 100 % for each problem instance. The BBO-selection algorithm reduces the size of original feature vector to 52%, 50%, 50.7%, and 50% for problem instance of 20X20, 30X30, 40X40, and 50X50 respectively. For example, if the DCT of an image is calculated and 20X20 DCT subset is taken from upper left of DCT array, there are total 400 features which are given as an input to BBO-FS algorithm. BBO-FS reduces the 400 features to 219 which means only 219 features are required to recognize the face from facial gallery.

IV. CONCLUSION

In this paper, a novel BBO-based feature selection algorithm for FR is proposed. The algorithm is applied to feature vectors extracted by Discrete Cosine Transform. The algorithm is utilized to search the feature space for the optimal feature subset. Evolution is driven by a fitness function defined in terms of class separation. The classifier performance and the length of selected feature vector were considered for performance evaluation using the ORL face database. Experimental results show the superiority of the BBO-based feature selection algorithm in generating excellent recognition accuracy with the minimal set of selected features.

REFERENCES

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey", *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.
- [2] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.
- [3] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000.

- [4] M. A. Turk and A. P. Pentland, "Face Recognition using Eigenfaces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.
- [5] L. Du, Z. Jia, and L. Xue, "Human Face Recognition Based on Principal Component Analysis and Particle Swarm Optimization -BP Neural Network," *Proc 3rd Conference. on Natural Computation (ICNC 2007)*, vol. 3, pp. 287-291, August 2007.
- [6] X. Yi-qiong, L. Bi-cheng and W. Bo, "Face Recognition by Fast Independent Component Analysis and Genetic Algorithm," *Proc. of the 4th International Conference on Computer and Information Technology (CIT'04)*, pp. 194-198, Sept. 2004.
- [7] X. Fan and B. Verma, "Face recognition: a new feature selection and classification technique," *Proc. 7th Asia-Pacific Conference on Complex Systems*, December 2004.
- [8] A. S. Samra, S. E. Gad Allah, R. M. Ibrahim, "Face Recognition Using Wavelet Transform, Fast Fourier Transform and Discrete Cosine Transform," *Proc. 46th IEEE International Midwest Symp. Circuits and Systems (MWSCAS'03)*, vol. 1, pp. 272- 275, 2003.
- [9] A. S. Samra, S. E. Gad Allah, R. M. Ibrahim, "Face Recognition Using Wavelet Transform, Fast Fourier Transform and Discrete Cosine Transform," *Proc. 46th IEEE International Midwest Symp. Circuits and Systems (MWSCAS'03)*, vol. 1, pp. 272- 275, 2003.
- [10] Z. Yankun and L. Chongqing, "Efficient Face Recognition Method based on DCT and LDA", *Journal of Systems Engineering and Electronics*, vol. 15, no. 2, pp. 211-216, 2004.
- [11] C. Podilchuk and X. Zhang, "Face Recognition Using DCT-Based Feature Vectors," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, vol. 4, pp. 2144-2147, May 1996.
- [12] F. M. Matos, L. V. Batista, and J. Poel, "Face Recognition Using DCT Coefficients Selection," *Proc. of the 2008 ACM Symposium on Applied Computing, (SAC'08)*, pp. 1753-1757, March 2008.
- [13] M. Yu, G. Yan, and Q.-W. Zhu, "New Face recognition Method Based on DWT/DCT Combined Feature Selection," *Proc. 5th International Conference on Machine Learning and Cybernetics*, pp. 3233-3236, August 2006.
- [14] Z. Pan and H. Bolouri, "High Speed Face Recognition Based on Discrete Cosine Transform and Neural Networks," Technical Report, Science and Technology Research Center (STRC), University of Hertfordshire.
- [15] Z. M. Hafed and M. D. Levine, "Face Recognition Using Discrete Cosine Transform", *International Journal of Computer Vision*, vol. 43, no. 3, pp. 167-188, 2001.
- [16] X. Fan and B. Verma, "Face recognition: a new feature selection and classification technique," *Proc. 7th Asia-Pacific Conference on Complex Systems*, December 2004.
- [17] D.-S. Kim, I.-J. Jeon, S.-Y. Lee, P.-K. Rhee, and D.-J. Chung, "Embedded Face Recognition based on Fast Genetic Algorithm for Intelligent Digital Photography," *IEEE Trans. Consumer Electronics*, vol. 52, no. 3, pp. 726-734, August 2006.
- [18] M. L. Raymer, W. F. Punch, E. D. Goodman, L.A. Kuhn, and A. K Jain, "Dimensionality Reduction Using Genetic Algorithms," *IEEE Trans. Evolutionary Computation*, vol. 4, no. 2, pp. 164-171, July 2000.
- [19] H. R. Kanan, K. Faez, and M. Hosseinzadeh, "Face Recognition System Using Ant Colony Optimization-Based Selected Features," *Proc. IEEE Symp. Computational Intelligence in Security and Defense Applications (CISDA 2007)*, pp 57-62, April 2007.
- [20] D. Simon, "Biogeography-based optimization", *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702-713, 2008.
- [21] D. Whitley, S. Rana, R.B. Heckendorn, "The island model genetic algorithm: on separability, population size and convergence", *Journal of Computing and Information Technology* 7 (1998) 33-47.
- [22] Gong, W., Cai, Z., Ling, C.X., Li, H., "A real-coded biogeography-based optimization with mutation", *Applied Mathematics and Computation*, vol. 216, no. 9, pp. 2749-2758, 2011.
- [23] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000.
- [24] F. M. Matos, L. V. Batista, and J. Poel, "Face Recognition Using DCT Coefficients Selection," *Proc. of the 2008 ACM Symposium on Applied Computing, (SAC'08)*, pp. 1753-1757, March 2008.
- [25] H. R. Kanan, K. Faez, and M. Hosseinzadeh, "Face Recognition System Using Ant Colony Optimization-Based Selected Features," *Proc. IEEE Symp. Computational Intelligence in Security and Defense Applications (CISDA 2007)*, pp 57-62, April 2007.

AUTHORS PROFILE



Navdeep Kaur has done B-Tech (Hons.) in Computer Science & Engineering & scored 81% marks from Punjab Technical University, Jalandhar (India) in 2005 and MTech in Computer Science & Engineering from Guru Nanak Dev Engineering College, Ludhiana of India in 2009. She is currently working as a lecturer in computer science and IT department of Rayat institute of Engineering & Information Technology of India.



Er. Poonam Gupta has done BTech (First Division) in Information Technology & scored 67.64% marks from Kurukshetra University, Kurukshetra (India) in 2007 and M.Tech in Computer Science & Engineering from Rayat institute of Engineering & Information Technology, Railmajra of India in 2011. She is currently working as a lecturer in computer science and IT department of Rayat Polytechnic college (Evening shift) of Punjab, India. She has presented 3 papers in National Conferences.



Amanpreet Kaur has done B.Tech in Computer Science & Engineering and scored 73% marks from Punjab Technical University, Jalandhar (India) in 2007 and M.Tech in also the same stream from Rayat Institute of Engineering & Information Technology, Railmajra, Punjab, India in 2010. She scored 71% marks in her post graduation. She is working as a lecturer in Computer Science & I.T. department in Rayat Institute of Engineering & information technology, Railmajra, Punjab, India.

A Novel Steganographic Methodology For Secure Transmission Of Images

B.V.Ramadevi*D.LalithaBhaskari*P.S.Avadhani*

Research ScholarAssociate ProfessorProfessor

bvramadevi@yahoo.comlalithabhaskari@yahoo.co.inpsavadhani@yahoo.com

*Department of Computer Science & Systems Engineering
AUCE(A), Andhra University, Visakhapatnam

Abstract :

In recent days steganographic techniques have gained a lot of significance in many of the security applications. In this paper a two layered secure methodology for transmitting multimedia data is proposed and implemented. In the first layer, encoding based compression of the message to be hidden is done based on G₀delization and Alphabetic coding(AC). In the second layer a steganographic approach is adopted for embedding of the encoded text into the cover image under frequency domain and the obtained stego image is transmitted securely using a novel encryption and decryption methods.

Keywords: Alphabetic Coding, Cover image, Encryption, Decryption, G₀delization, stego image.

1. Introduction

Steganography is a modern and dynamically developing part of information security which protects information by its hiding techniques. Technically speaking, steganography is a covert communication technology, which allows secret information to be embedded into a cover/host message without significantly damaging the content of the cover message. The message usually will be an image and the secret

information which is to be embedded is called the stego message. Today steganography has got many different ways to hide information like images, audio, video etc. In fact steganography, cryptography, watermarking are all different branches of information hiding. Information hiding also often termed as data hiding or multimedia data hiding is a term covering a wide range of problems beyond embedding messages in content. The term hiding can refer to either making the information perceptible or keeping the existence of information secret. Compared with text or binary data, multimedia data often has high redundancy, large volumes and real-time operations. All these properties require that multimedia data should be compressed, encrypted and securely transmitted for the required applications. During the past decades, various multimedia encryption algorithms have been proposed and studied. They can be classified into three types which are image encryption, audio encryption, and video encryption. Generally, for different content, different encryption algorithms should be adopted[1]. In this paper a novel image encoding based on G₀delization[2], compression based on alphabetic coding[2], embedding the data into the cover image using a steganographic method termed as middle band coefficient exchange

algorithm[3] and then secure transmission of the data using a methodology based on encryption and decryption[4]are proposed and presented.

2. Basic Concepts

In this paper, the proposed methodology is divided into two layers(modules)for embedding and securely transmitting the multimedia data. The first layer consists of converting(encoding) the secret image into string of Gödel Number Sequence throughGödelization, compressing the encoded string using alphabetic coding andembedding the encoded compressed stringinto the cover image using middle band coefficient exchange method[3] under frequency domain.In the second layer, encryption and decryption techniquesare implemented for secure transmission of the data. In this section a brief description of the definitions and the concepts are presented.

In the first layer, the secret image which is to be transmitted securely is converted into Gödel Number Sequence(GNS) through the concept of Gödelization[2]. According to it,the intensity values at a point $f(x,y)$ in the image are transformed into the power of its primes. Consider a pixel value 39 which can be factorized as $3^1 \times 13^1$. So the Gödel number sequence of $39 = GN(0,1,0,0,0,1)$. The sequence 0,1,0,0,0,1 can be encoded as $3^1 \times 13^1$ as $GN(0)=2, GN(1)=3, GN(2)=5$ and so on. After converting each and every pixel into the corresponding GNS, alphabetic compression technique(AC)[2] is applied to compress the GNS. According to AC, if the GNS has a sequence of more 0's and 1's, we represent 0's with 'A', 1's with 'B', 2's with 'C' and so on. If we encounter more than 3 same characters then, the number of occurrences are represented first followed by the character. After applying AC, there is a considerable amount of compression achieved. The obtained encoded compressed string is embedded into the cover image using middle band coefficient exchange[2] method under frequency domain. According to literature survey, embedding of secret data into the digital images can be done in two domains. One is

spatial domain[4] where the intensity values(pixels) of the image are manipulated and data is hidden in the intensity values of the images. The second method is frequency domain where the frequency components of the digital images are considered[5,6]. The secret data is embedded into the frequency components of the image. It is observed that spatial domain manipulations are easy when compared to frequency domain, yet frequency domain provides more security when compared to spatial domain techniques. So in this work, frequency domain is chosen as the media and discrete cosine transforms(DCT)[7] are considered. After embedding the data into the cover image, a key(k) and a stego image is obtained which is given as input to the second layer. In the second layer, the obtained Key and the stego image are encrypted and transmitted. At the decoding end, the decoder decrypts the stego image with his private key.

3. Encryption and Decryption

With the continuing development of both computer and Internet technology, multimedia data (images, videos, audios, etc.) is being used more and more widely, in applications such as video conferencing, broadcasting, education, commerce, politics etc., and so the security concerns are also increasing. To maintain security, multimedia data should be protected before transmission or distribution. The typical protection method is the encryption technique [9] which transforms the data from the original form into an unintelligible form. Until now, various data encryption algorithms have been proposed and widely used, such as AES, RSA, or IDEA [9,10], most of which are used in text or binary data. It is difficult to use them directly in multimedia data, for multimedia data are often of high redundancy, of large volumes and require real-time interactions, such as displaying, cutting, copying, bit rate conversion, etc. So in this paper, the multimedia data(the secret image) is encoded and compressed using Gödelization and alphabetic coding as explained in the previous section. This encoded compressed string is embedded into the cover image under frequency domain using middle band coefficient exchange method[10] to obtain a key and a stego

image. The key and stego image combined together are encrypted and transmitted. At the decoding end, it is decrypted using the decoder's private key. The encryption and decryption models are shown below in fig 1 and 2 respectively.

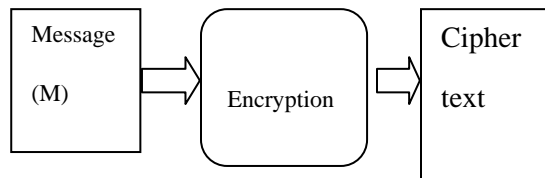


Fig 1. Scheme for Encryption

Here the stego image and the key generated are given as input to the encryption algorithm. The cipher text is given as input to the decoder which decodes the cipher text along with the receiver's private key. The decoding process is as shown below.

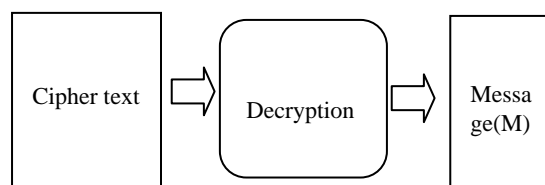


Fig. 2 Scheme for decryption

4. Proposed Methodology

The proposed methodology provides two layered security when compared to the traditional methods. The whole scheme of the proposed methodology can be viewed as shown below in figures 3 and 4. The secret image (data) which is to be hidden is encoded into Gödel

Number string (GNS) using Gödelization technique, later the encoded string is compressed using AC technique. This encoded compressed string is embedded into the cover image using middle band frequency exchange method in the frequency domain. The output obtained after embedding the data is a key and a stego image. This provides the first layer of security. The obtained key and the stego image is now encrypted and decrypted. In this model, MD5 is used for hashing, IDEA is used for the encryption process. This provides second layer of security and the data is transmitted. At the decoder end the data is decrypted so as to obtain the message which is the key (k) and the image data (M) which is obtained during embedding under frequency domain. This data is decoded with the key, then reverse Gödelization is used to obtain GNS, upon which reverse alphabetic coding is applied to obtain the image data which is reconstructed to obtain the secret image.

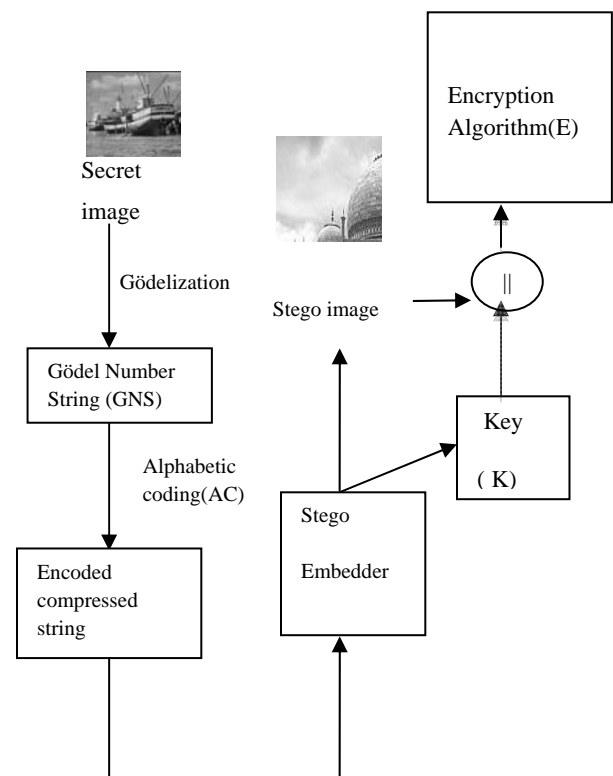


Fig 3. Scheme of proposed method at the encryption side

The decryption side scenario is given below in fig 4. After decoding image can be reconstructed at the receiver's side from the obtained data.

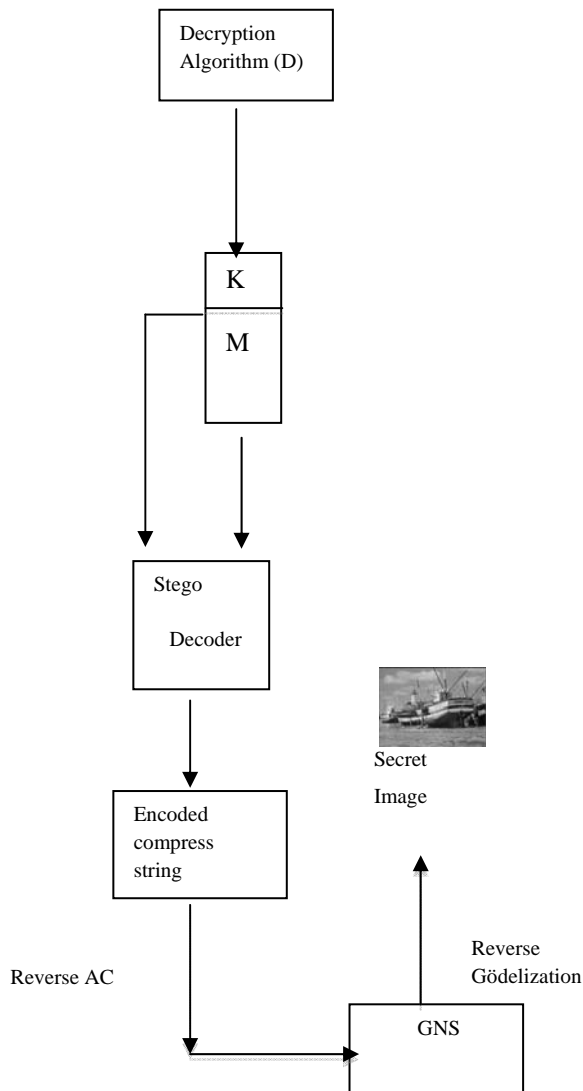
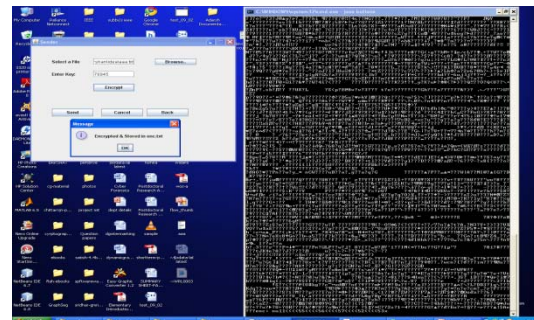


Fig 4. Scheme of proposed method at the decryption side.

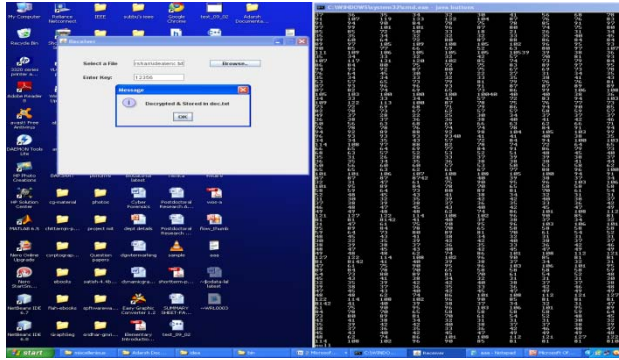
5. Results

The proposed model is being implemented using Matlab 7.0 and JAVA. The results proved to be more secure and satisfactory. Some test cases are provided here.

Sender's side encryption:



Receiver's side decryption:



6. Conclusions & Future Work

Thus by implementing the proposed model it is observed that not only the data payload capacity has increased but security is also enhanced when compared to other methods. By using Gödelization data is being encoded which is in turn compressed so as to increase the data payload capacity. This data is embedded into the cover image using middle band frequency exchange methods under frequency domain which is proved to be secure. This data is again transmitted using a novel encryption/decryption model and data is retrieved at the decoder end. It is also observed that, after embedding the data into the cover image there is no perceptual difference between the cover image and the stegoimage which satisfies the property of a blind steganographic scheme. In a blind scheme the cover image is not required at the decoder's end and the implementation results are satisfactory. As future enhancement, the same technique can be enhanced for RGB images so

that more data payload capacity can be achieved which sometimes is necessary in some applications.

7. References:

- [1] S. Lian, J. Sun, D. Zhang, and Z. Wang, "A Selective Image Encryption Scheme Based on JPEG2000 Codec", 2004, Pacific-Rim Conference on Multimedia (PCM2004), Lecture Notes in Computer Science, Vol. 3332, pp 65–72.
- [2] B.V.Rama Devi, P.PrapoornaRaja, D.LalithaBhaskari, P.S.Avadhani, "A New Encryption Method for Secure Transmission of Images", 2010, International Journal of Computer Science and Engineering (IJCSE), Vol.02, No.09, pp 2801-2804.
- [3] B.Neil F. Johnson., Stefan C.Katzenbeisser, "Chapter 3: A Survey of Steganographic Techniques", Arctech House, London, 2000.
- [4] P.S.Avadhani, D.LalithaBhaskari, "A Blind Scheme Watermarking Technique Using Gödelization Technique for RGB images under spatial domain", International Conference on i-warfare (ICIW-2010), Dayton, USA, 8th -9th April 2010, pp 373-377.
- [5] El-Gayyar, Joachim von zurGathen, "Watermarking Techniques Spatial Domain", Digital Rights Seminar, University of Bonn, Germany, May 2006.
- [6] D. Tzovaras, N. Karagiannis, M. G. Strintzis, "Robust image watermarking in the subband or discrete cosine transform domain", 9th European Signal Processing Conference (EUSIPCO'98), Greece, 8–11, Sept. 1998, pp 2285–2288.
- [7] Bret Dunbar, "A detailed look at Steganographic Techniques and their use in an Open-Systems Environment", The Information Security Reading Room, SANS Institute, 2002.

[8]<http://www.mishmash.com/fredspgp/pgp.htm>

[9]S. A. Vanstone, A. J. Menezes, and P. C. Oorschot, 1996,“*Handbook of Applied Cryptography*”,Boca Raton, FL: CRC Press.

[10]B.V.Rama Devi et.al., “ A New Steganographic Algorithm for Image Hiding using Gödelization under frequency domain”,2011, IJARCS, Vol 2,No.2, pp 519-522.

A New Dynamic Data Allocation Algorithm for Distributed Database

Fardin Esmaili Sangari
Sama Technical and vocational training college
Islamic Azad university
Urmia branch, Urmia, Iran
Fardin_e_s@yahoo.com

Seyed Mostafa Mansourfar
Sama Technical and vocational training college
Islamic Azad university
Sahand branch, Sahand, Iran
Mostafa.mansourfar@gmail.com

Abstract— *Data and fragment allocation is an important issue in distributed database systems. Data allocation is carried out based on data access dynamic and static patterns. This paper proposes a new strategy for data allocation named Relative Threshold Algorithm (RTA) in non-repeated distributed database. Proposed algorithm does reallocation data fragments by changing access pattern to data fragments. This algorithm proposes data fragments migrate at the site that has at most availability to fragments. Simulation results show that RTA performance is better than existing algorithms in term of hit ratio. It also reduces requirement space. We believe the reduction of storage overhead make RTA more attractive in distributed database systems.*

Keywords-component: Distribute Database; Dynamic Data Allocation

I. Introduction

Database and network technologies have been the most important problems in creating distributed database systems, for the past decade. A distributed database system is consists of a collection of sites connected communication network, in which each site is a database system in its own right, but the sites have agreed to work together, so a user at any site can access data anywhere in the network exactly as if the data were all stored at the user's own site [1]. Distributed database systems use data allocation for achieving two aims. First is total data transmission cost minimized for process (i.e., the maximum number of fragments that can be allocated in a site) and Second one is the unifying of implementation strategy. The majority concern of a distributed database system is the designing of the fragmentation and allocation of the underlying data. Fragmentation unit can be a file where allocation issue becomes the file allocation problem [2]. However, data allocation is a NP-complete problem [3]. So, quick allocation requires creation of efficient solution. Moreover, optimal allocation of database hardly is employed by a distributed database system on query strategy.

A few papers have been recently proposed for data allocation problem. Chu in [4] has considered this problem. Repetitive and non-repetitive models conducted in [5][6] and [7][8] address issue dynamic file allocation. In [6][7][8] and [9] have been presented various solutions for data allocation

in distributed systems. These papers have been performed data allocation depending on static data access patterns or query access patterns. Access probability of nodes to data fragmentations is stable in static environment. While these changes in dynamic environments and using of static methods frequently reduces database performance. Dynamic algorithm has been presented for data allocation in non-replicate database systems called threshold algorithm [7]. Threshold algorithm transfers data fragmentation among sites according to change data access pattern. It focuses on load balance. This algorithm provides data allocation with low hit ratio. In other words, the requirement probability of that site is low to fragment in site and it doesn't completely consider number of other sites access while takes into account and only the last site has access to data during data transfer to other sites. We aim to focus on the disadvantages and we attempt to eliminate them. The rest of the paper is as follows. In section 2, we review threshold algorithm. Proposed algorithm is presented in section 3. In Section 4, simulation results of proposed algorithm have been showed. Finally, section 5 is the conclusion.

II. Threshold algorithm

Threshold algorithm is one of the dynamic allocation algorithms which transfer data fragments among sites according to changing patterns [7][10][11]. Threshold algorithm stores only one counter for each fragment. Figure 1 shows fragment i with its associated counter.

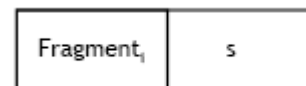


Figure 1. Any fragment i in threshold algorithm

In the threshold algorithm, the initial value of the counter is zero. The counter value is increased by one for each remote access to the fragment. It is reset to zero for a local access. Whenever the counter exceeds a predetermined threshold value, the ownership of the fragment is transferred to another node. At this point, the critical question is, which node will be the new owner of the fragment? The algorithm

gives very little information about the past accesses to the fragment. In fact, throughout the entire access history only the last node which accessed the fragment is known. Two strategies have been selected for current possessor. Whether new possessor is selected randomly, or last accessing node is selected as new possessor. In initial strategy, the randomly chosen node could be one that has never accessed the fragment before. Therefore, latter strategy heuristically is better. Initially all fragments are distributed to the nodes randomly. A threshold value is set by δ . Every node j , threshold algorithm executes for every fragment i that have been stored. It reduces traffic two nodes which have threshold value exceed one ($\delta > 1$). One of the important problems in threshold algorithm is the exact choice of threshold value. Because of this, value affects on fragments movement (mobility of the fragments) directly. If threshold value increases, fragment will tend to remain more in current node. Otherwise, as the threshold value decreases, fragment tendency will visit more sites.

In threshold algorithm, if n fragments are in a site then n distinctive counter are requirement. If site B consecutively accesses to fragment in site A then counter increases by one and counter is tended to threshold value. Now, if site A randomly accesses to fragment that before site B consecutively accesses it then counter be zero.

If site B consecutively accesses to fragment in site A and site C accesses to this fragment for first time and with this access, counter value equal with threshold value then fragment is transferred to site C because site C has performed last access. This events increase response time.

III. Proposed Algorithm

Our proposed algorithm uses two fields for every site. Number of fields doesn't depend on fragments number which resides in site. One of fields count number of accesses and other shows last fragment which has access to current site. The fragment tends to stay at the node with higher access probability. As the access probability of the node increases, the tendency to remain at this node also increases. It is also shown that as the threshold value increases, the fragment will tend to stay more at the node with higher access probability. At every access, name of fragment is compared with counter if they are similar counter increased by one. Counter is set to zero when site accesses to fragment for first time and then the name of fragment is recorded in identifier field. Our algorithm computes total number of accesses whether these accesses are local or remote. It is important that the number of accesses is interval. This algorithm increases probability of fragment resident in site. However, response time decreases, because it doesn't require any information replacement from remote site. Threshold algorithm is a centralized algorithm. If site failed, total site information would waste. Our proposed algorithm is

distributed. We eliminate single point of failure. If that site crashed, other sites access to information yet is there and only crashed site information will be destroyed. Our proposed algorithm raise hit ratio. It reduces data replacement due to locality. This would be show as follow.

We make our work assumptions as follow.

- Initially, fragments are randomly distributed in the sites.
- Initially, counter value is zero
- An incremental counter is used. The initial value of the counter is zero.
- if the name of access fragment is same as the name of identifier field then For each access to fragment, counter value increases by one

Fragment	counter
----------	---------

Figure2. The fragment in each site at proposed algorithm

Relative threshold algorithm:

- Step 1. Initial counter value is set zero for all sites and distribute fragments randomly between sites. (at each site counter=0)
- Step 2. Process the access request for stored fragment.
- Step 3. For each request (locally or remote), counter value increase one, if the access is repetitive. go to step 2.
- Step 4. If name of requested fragment is not same as the fragment field, set counter by zero is replaced identifier field with new fragment name.
- Step 5. If counter value exceeds threshold value ($\text{counter} > \delta$) and fragment is in site then counter will be zero else fragment is transferred to access site and counter will be zero.
- Step 6. Refer to step 2.

We suppose sites topology as in figure 3. Site 2 wants to access fragment of site1, so it increased one to counter and fragment field value become equal to A. each sequential access increases counter value, if site 2 finds existent data in A. if this value is higher threshold value, data will move to site 2. If site 2 accesses to data unlike A, counter value will be zero. And fragment field value will be replaced by a new fragment name.

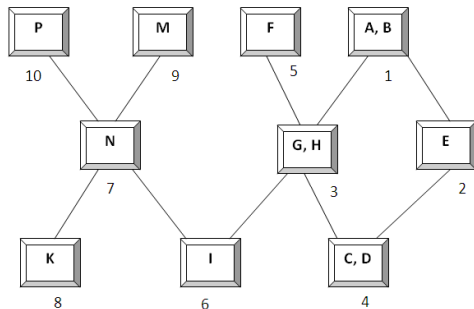


Figure3. Example of sites topology

IV. Simulation Results

In this section, we evaluate the proposed algorithm and compare it with threshold algorithm and show our algorithm which has better performance. In this simulation, the number of fragment is between 100 and 9000. Initially, these fragments are randomly distributed between sites. Experiments were examined in different environments.

In first scenario, we consider number of sites variably and assume threshold value as stable (figure 2).

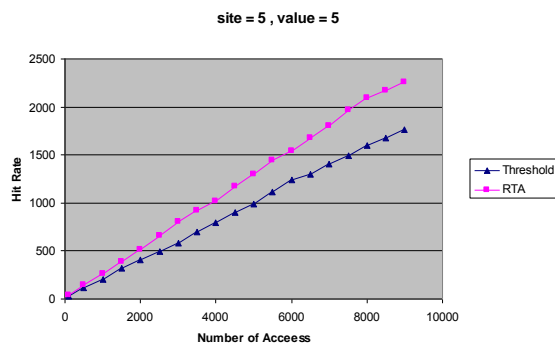


Figure 4. Compare of Hit Rate in RTA & Threshold algorithm with different access numbers

In this experiment, hit ratio factor of data fragment length is 2500, threshold value is 5 and number of sites is 5. So simulation results in figure 3 show proposed algorithm increases fragment hit ratio when requested fragment exist in current site.

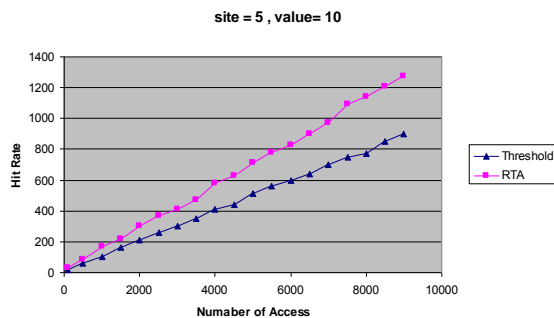


Figure 5. compare of Hit Rate in RtA & Threshold algorithm with different access numbers

Experiment is repeated with number of site 5 and threshold 10 and similar results have almost been achieved. Whatever environment be more intense, higher hit ratio would be achieved.

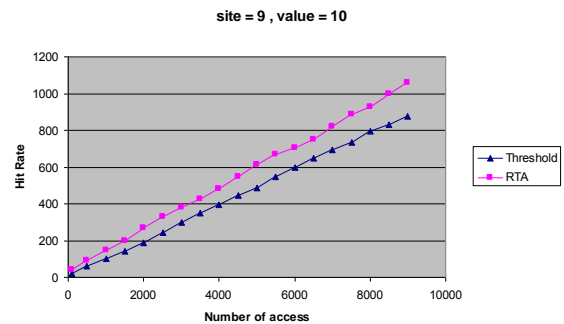


Figure 6. compare of Hit Rate in RTA & Threshold algorithm with different access numbers and different sites & value numbers

V. Conclusion

In this article we introduce a new method to distributed data fragment of Distributed Database System. RTA is based on threshold algorithm that uses different strategy for data transmission. In our experiments, we consider hit ratio. This simulation is configurable for testing different network topologies and different data request and/or allocation conditions. Result of experiment shows the RTA hit rate is better than threshold algorithm and achieve better improvement of threshold algorithm. We use non-repeated distributed algorithm. In future, we can consider RTA in repeated distributed algorithm.

References

- [1] Baseda, S. Tasharofi, M. Rahgozar, "Near Neighborhood Allocation: A Novel Dynamic Data Allocation Algorithm in DDB", CSICC 2006.
- [2] Navathe, S.B., S. Ceri, G. Wiederhold and J.Dou, "Vertical Partitioning Algorithms for Database Design", ACM Transaction on Database Systems, 1984
- [3] Y. F. Huang and J. H. Chen, "Fragment Allocation in Distributed Database Design", Journal of Information Science and Engineering 17, 491-506, 2001
- [4] Ahmad, I, K. Karlapalem, Y. K. Kwok and S. K. "Evolutionary Algorithms for Allocating Data in Distributed Database Systems", International Journal of Distributed and Parallel Databases, 11: 5-32, The Netherlands, 2002.
- [5] A. Brunstroml, S. T. Leutenegger and R. Simhal, "Experimental Evaluation of Dynamic Data Allocation Strategies in a Distributed Database with changing Workloads", ACM Transactions on Database Systems, 1995
- [6] A. G. Chin, "Incremental Data Allocation and ReAllocation in Distributed Database Systems", Journal of Database Management; Jan-Mar 2001; 12, 1; ABI/INFORM Global pg. 35
- [7] T. Ulus and M. Uysal, "Heuristic Approach to Dynamic Data Allocation in Distributed Database Systems", Pakistan Journal of Information and Technology 2 (3): 231-239, 2003
- [8] S. Voulgaris, M.V. Steen, A. Baggio, and G. Ballintjn, "Transparent Data Relocation in Highly Available Distributed Systems". Studia Informatica Universalis. 2002

- [9] L. C. John, "A Generic Algorithm for Fragment Allocation in Distributed Database Systems", ACM, 1994
- [10] Basseda. R, "Fragment Allocation in Distributed Database Systems", Database Research Group, 2006
- [11] Basseda. R, "Data Allocation In Distributed Database Systems", Technical Report No. DBRG. RB-ST. A50715, 2005.

Establishing an Effective Combat Strategy for Prevalent Cyber-Attacks

Vivian Ogochukwu Nwaocha
University of Nigeria, Nsukka
Computer Science Department
ogochukwuvee@gmail.com

Inyiama H.C.
University of Nigeria, Nsukka
Computer Science Department
drhcinyiama@gmail.com

Abstract—As organisations continue to incorporate the Internet as a key component of their operations, the global cyber-threat level is increasing. One of the most common types of cyber-threats is known as the Distributed Denial of Service (DDoS) attack – an attack preventing users from accessing a system for a period of time. Recent DDoS attacks have left large corporate and government networks inaccessible to customers, partners and users for hours or days, resulting in significant financial, reputational, and other losses. The attack power of a Distributed DoS (DDoS) attack is based on the massive number of attack sources instead of the vulnerabilities of one particular protocol. DDoS attacks, which aim at overwhelming a target server with an immense volume of useless traffic from distributed and coordinated attack sources, are a major threat to the stability of the Internet. The number and assortment of both the attacks as well as the defense mechanisms is outrageous. Though an array of schemes has been proposed for the detection of the presence of these attacks, classification of the TCP flows as a normal flow or a malicious one, identifying the sources of the attacks and mitigating the effects of the attacks once they have been detected, there is still a dearth of complete frameworks that encompass multiple stages of the process of defense against DDoS attacks. The growing use of cloud computing services and shared infrastructure is further increasing the importance of having a considered plan for managing such attacks. For a proactive mitigation against DDoS attacks, we propose an integrated framework which would handle the classification, mitigation and traceback of these attacks. Thus, developing an effective mitigation strategy is an important measure to minimize the risk posed to an organisation by the threat of DDoS attacks.

Keywords—attacks; classification; cyber, detection; distributed denial of service (DDoS); intrusion; mitigation, traceback;

I. INTRODUCTION

The growing population using public network has brought about an increase in the incidence of network intrusion. Hence the need for an equivalent increase in business owner's duty to guarantee due diligence and fiduciary responsibility with respect to protecting users against all causes of loss or damage. The potential costs of failing to do so can in fact be quite enormous. Amongst the security threats, the most severe to the steady functioning of any network are Distributed Denial-of-Service (DDoS) attacks. Distributed Denial of

Service (DDoS) is one of the major threats for the Internet because of its ability to create a huge volume of unwanted traffic [1]. The primary goal of these attacks is to prevent access to a particular resource such as a Web site [2].

The first reported large-scale DDoS attack occurred in August, 1999, against the University of Minnesota [3]. This attack shut down the victim's network for more than two days. In the year 2000, a DDoS attack stopped several major commercial Web sites, including Yahoo and CNN, from performing their normal activities [3]. In [4], D. Moore *et al.* used backscatter analysis on three week-long datasets to assess the number, duration and focus of DDoS attacks, and to characterize their behaviour. They found that more than 12,000 attacks had occurred against more than 5,000 distinct victims in February, 2001. In October, 2002, the Domain Name Systems (DNS) in the Cooperative Association for Internet Data Analysis (CAIDA) network became the victim of a heavy DDoS attack. Many legitimate users could not access web sites because their DNS requests were not able to reach root DNS servers. The congestion caused by the DDoS attack forced routers to drop these requests [5]. A more serious DNS-based DDoS attack was reported in March, 2006 [6]. Instead of attacking DNS servers directly, this new type of DDoS attack just used DNS servers as reflectors to create a stronger attack. This kind of DDoS is harder to be stopped than normal DDoS attacks due to complicated DNS protocols and interaction among multiple DNS servers. During two months, 1,500 individual Internet protocol addresses were attacked using this approach.

As organisations continue to incorporate the Internet as a key component of their operations, the global cyber-threat level is increasing. One of the most common types of cyber-threats is known as the Distributed Denial of Service (DDoS) attack – an attack preventing users from accessing a system for a period of time. Recent DDoS attacks have left large corporate and government networks inaccessible to customers, partners and users for hours or days, resulting in significant financial, reputational, and other losses. The attack power of a Distributed DoS (DDoS) attack is based on the massive number of attack sources instead of the vulnerabilities of one

particular protocol. DDoS attacks, which aim at overwhelming a target server with an immense volume of useless traffic from distributed and coordinated attack sources, are a major threat to the stability of the Internet. The number and assortment of both the attacks as well as the defense mechanisms is outrageous. Though an array of schemes has been proposed for the detection of the presence of these attacks, classification of the TCP flows as a normal flow or a malicious one, identifying the sources of the attacks and mitigating the effects of the attacks once they have been detected, there is still a dearth of complete frameworks that encompass multiple stages of the process of defense against DDoS attacks. The growing use of cloud computing services and shared infrastructure is further increasing the importance of having a considered plan for managing such attacks. For a proactive mitigation against DDoS attacks, we propose an integrated framework which would handle the classification, mitigation and traceback of these attacks. Thus, developing an effective mitigation strategy is an important measure to minimize the risk posed to an organisation by the threat of DDoS attacks.

II. DISTRIBUTED DENIAL OF SERVICE ATTACKS

A **Denial of Service (DoS)** attack is commonly characterized as an event in which a legitimate user or organisation is deprived of certain services such as e-mail or network connectivity, that they would normally expect to have. DoS attacks [7, 8] inject maliciously-designed packets into the network to deplete some or all of these resources. The attack power of a Distributed DoS (DDoS) attack [9] is based on the massive number of attack sources instead of the vulnerabilities of one particular protocol. DDoS attacks, which aim at overwhelming a target server with an immense volume of useless traffic from distributed and coordinated attack sources, are a major threat to the stability of the Internet. The number and assortment of both the attacks as well as the defense mechanisms is outrageous. Though an array of schemes has been proposed for the detection of the presence of these attacks, characterizing of the flows as a normal flow or a malicious one, identifying the sources of the attacks and mitigating the effects of the attacks once they have been detected, there is still a dearth of complete frameworks that encompass multiple stages of the process of defense against DoS attacks.

For a proactive mitigation against flooding-based DDoS attacks, we propose an integrated framework which would handle the classification. As one of the major security problems in the Internet, a denial-of-service (DoS) attack always attempts to stop the victim from serving legitimate users. A distributed denial-of-service (DDoS) attack is a DoS attack which relies on multiple compromised hosts in the network to attack the victim. There are two types of DDoS attacks. The first type of DDoS attack has the aim of attacking the victim to force it out of service for legitimate users by exploiting software and protocol vulnerabilities of the system [10]. The second type of DDoS attack is based on a huge volume of attack traffic, which is known as a flooding-based

DDoS attack. A flooding-based DDoS attack attempts to congest the victim's network bandwidth with real-looking but unwanted IP data. As a result, legitimate IP packets cannot reach the victim due to a lack of bandwidth resource. To amplify the effects and hide real attackers, DDoS attacks can be run in two different distributed coordinated fashions. In the first one, the attacker compromises a number of agents and manipulates the agents to send attack traffic to the victim. The second method makes it even harder to determine the attack sources because it uses reflectors. A reflector is any host that will return a packet if it receives a request packet [11]. For example, a Web server can be a reflector because it will return a HTTP response packet after receiving a HTTP request packet. The attacker sends request packets to servers and fakes victim's address as the source address. Therefore, the servers will send back the response packets to the real victim. If the number of reflectors is large enough, the victim network will suffer exceptional traffic congestion. Before we introduce the DDoS attack architectures and mechanisms, we give two basic definitions. First, the DDoS attack traffic is the traffic which is produced or triggered by the compromised agents. Second, the legitimate traffic is the traffic which is produced by the normal hosts. In order to analyze DDoS attacks, two basic distributed architectures of flooding-based DDoS attacks and common IP spoofing techniques were employed. Furthermore, we specify the basic mechanism of spoofing-based DDoS attacks and list three typical flooding-based DDoS attacks.

A. Distributed Cooperative Architecture of DDoS

Before real attack traffic reaches the victim, the attacker must cooperate with all its DDoS agents. Consequently, there must be control channels between the agents and the attacker. This collaboration requires that all agents send traffic based on commands received from the attacker. The network which consists of the attacker, agents, and control channels is called the attack networks. In [12], attack networks are divided into three types: the *agent-handle model*, the *Internet Relay Chat (IRC)-based model*, and the *reflector model*.

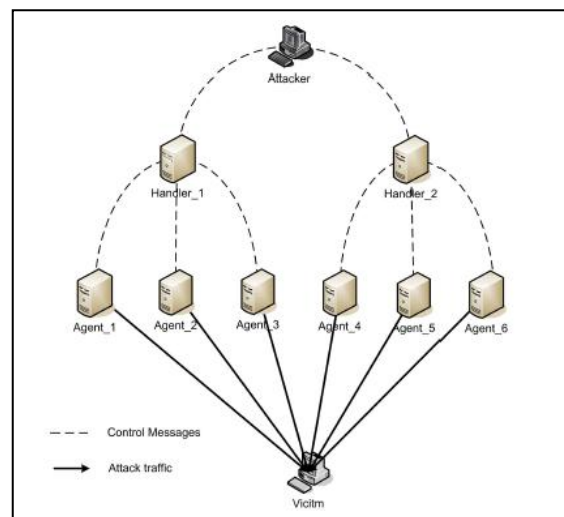


Figure 1 Classic Architecture of a DDoS Attack

The agent-handler model consists of three components: attacker, handlers, and agents. Figure 1 illustrates the typical architecture of the model. One attacker sends control messages to the previously compromised agents through a number of handlers, instructing them to produce unwanted traffic and send it to the victim. The architecture of IRC-based model is not that much different than that of the agent-handler model except that instead of communication between an attacker and agents based on handlers, an IRC communication channel is used to connect the attacker to agents [12]. Fig. 2. illustrates the architecture of an attack network in the reflector model. The reflector layer makes a major difference from the typical DDoS attack architecture. In the request messages, the agents modify the source address field in the IP header using the victim's address to replace the real agents' addresses. Then, the reflectors will in turn generate response messages to the victim. As a result, the flooding traffic which reaches the victim is not from a few hundred agents, but from a million reflectors [11]. An exceedingly diffused reflector-based DDoS attack raises the bar for tracing out the real attacker by hiding the attacker behind a large number of reflectors. Unlike some types of DDoS attacks, "the reflector does not need to serve as an amplifier" [11]. This means that reflectors still can serve other legitimate requests properly even when they are generating attack traffic. The attacker does not need to compromise reflectors to control their behaviours in the way that agents need to be compromised. Therefore, any host which will return a response if it receives a request can be a reflector. These features facilitate the attacker's task of launching an attack because it just needs to compromise a small number of agents and find a sufficient number of reflectors.

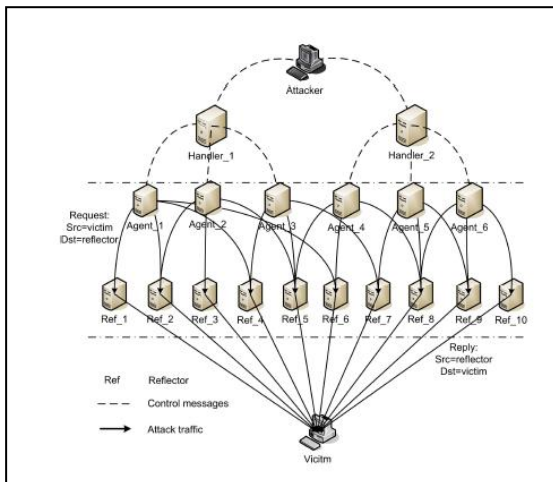


Figure 2. Architecture of a DDoS attack using reflectors

A. IP Spoofing

The template is used to format your paper and style the text. IP spoofing is used in all DDoS attacks as a basic mechanism to hide the real address of agents or the attacker. In a classical DDoS attack, the agents randomly spoof the source addresses in the IP header. In a reflector-based DDoS attack, agents must put the victim's address in the source address field. The spoofed addresses can be addresses of either existing or non-existing hosts. To avoid ingress filtering, the attacker can use addresses that are valid in the internal network because non-existing addresses have a high possibility of being filtered out. In the real-world, it is possible to launch an attack without IP spoofing if the attacker can compromise enough hosts. For this situation, the attacker would consider how to avoid to be traced out. Usually, the attacker will use a chain of compromised hosts. Tracing a chain which extends across multiple countries is very hard to be achieved. Furthermore, to compromise poorly monitored hosts in a network will make tracing more difficult due to a lack of information. In these situations, IP spoofing is not a necessary step for hiding the attacker.

C. Flooding DDoS Attack Mechanisms

Flooding-based DDoS attacks involve agents or reflectors sending a large volume of unwanted traffic to the victim. The victim will be out of service for legitimate traffic because its connection resources are used up. Common connection resources include bandwidth and connection control in the victim system. Generally, flooding-based DDoS attacks consist of two types: direct and reflector attacks [65]. Figure 3 is another view of the process of a direct flooding-based DDoS attack. The architecture of the direct attack is same as the typical DDoS attack reflected in Fig. 1

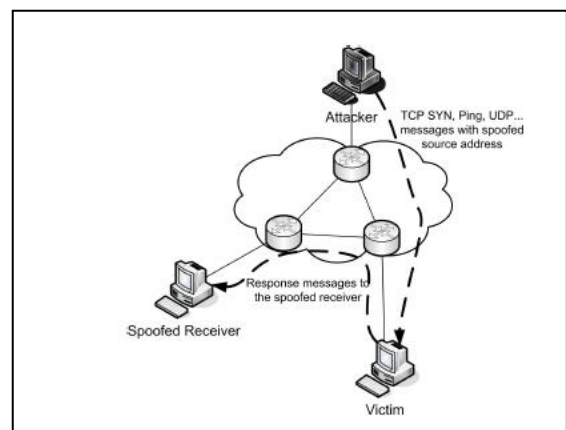


Figure 3. A Direct-Flooding Based DDoS Attack

The agents send the Transmission Control Protocol/Internet Protocol (TCP), the Internet Control Message Protocol

(ICMP), the User Datagram Protocol (UDP), and other packets to the victim directly. The response packets from the victim will reach the spoofed receivers due to IP spoofing. In a reflector attack, presented in Fig. 2.4, the response packets from reflectors truly attack the victim. No response packets need be sent back to reflector from the victim. The key factors to accomplishing a reflector attack include: setting the victim address in the source field of the IP header and finding enough reflectors. Basically, an attacker can utilize any protocol as the network layer platform for a flooding-based attack [10]. Direct attacks usually choose three mechanisms: TCP SYN flooding, ICMP echo flooding, and UDP data flooding [14]. The TCP SYN flooding mechanism is different from the other two mechanisms. It causes the victim to run out of all available TCP connection control resources by sending a large number of TCP SYN packets.

In a typical DDoS attack network, an attacker sends commands to compromised agents and requests that they send a large volume of traffic to overwhelm the bottleneck link in the victim network. To hide the attacker itself more deeply, a DDoS attack can construct an attack network with a reflector-based architecture. In the network, an attacker sends a packet whose source address has been set as the victim's address to reflectors.

III. RELATED WORK

Now we would review the existing combat strategies in this field, in order to compare our work with some associated work. Research in this area can be divided based on the following three issues: Classification, Mitigation and Traceback DDoS detection, DDoS response, and DDoS defense framework. The earliest work on DDoS defense led to the concept of network traceback [15] by Burch and Cheswick. Bellovin et.al. proposed ICMP-based out-of-band messaging in iTrace [16], while Snoeren et.al. proposed SPIE [17] employing packet logging, which was subsequently improved by Li et.al. in [18]. Belenky and Ansari proposed a deterministic packet marking scheme in [19], while Savage et.al. proposed a probabilistic packet marking (PPM) technique in [20], with subsequent enhancements made by others in [21] [22] [23] [24]. IP address fragmentation for efficient packet marking and their vulnerability to attacker induced noise have been studied in [25] and [26] respectively.

Recently, various encoding techniques have been used to progressively improve the performance of PPM schemes, as in Tabu marking [27], Local Topology marking [28], Space-Time encoding [29], Color Coding [30], and the use of Huffman Codes [31], Algebraic Geometric Codes [32] etc. Additionally various architectures for traceback have been explored, such as inter-domain traceback [33] and hybrid traceback [34] [35], in addition to some other radical approaches like in [36]. Research on mitigating DDoS attacks has proceeded in parallel, focusing on network ingress filtering [37], routing table enhancements as in SAVE [38], CenterTrack [39] and intelligent filtering [40]. The concept of

path fingerprints was exploited by Yaar et.al. in [41], and subsequently improved in [42]. Various other techniques involving path filtering [43] [44], statistical filtering [45] [46], and rate limiting [47] [48] have also been explored in literature. IP Marking [49] is traditionally used for IP Traceback.

The basic idea of the IP marking approach is that routers probabilistically write some encoding of partial path information into the packets during forwarding, so that based on this information the destination server can reconstruct the path that was taken by the packets. In [50], Song and Perrig have suggested Advanced and Authenticated Marking Schemes that encode the edge information in 16 bits of the packet to be marked. For this purpose, the 16-bit IP Identification field used for fragmentation in the IP header is overloaded, i.e. this field carries the encoding information instead of the regular packet fragmentation information. The obvious drawback in the methods discussed for IP Traceback is that they do not work for packets that are fragmented as the IP Identification field is overloaded for edge information.

Several methods have been proposed to characterize attack flows. In [51], a simple statistics-based mechanism to detect TCP SYN flood attacks was proposed. The idea is to detect deviation from an expected balanced SYN/FIN packet ratio using a nonparametric, cumulative sum method. However, such a simple technique is not foolproof as the attackers can mix their SYN and FIN packets. Subsequently, in [52] a spectral analysis method to distinguish attack flows from the normal ones by determining the periodicity in the packet process was proposed. But the method does so by using the Welch's modified periodogram, which has several disadvantages as compared to the EPSD technique used in this paper.

Bohacek [53], suggested a mitigating approach that relies on routers filtering enough packets so that the server is not overwhelmed while ensuring that as little filtering as possible is performed. He has proposed a solution wherein packets should be filtered at routers through which the attack packets are passing. But, it is a reactive mitigation technique that also has the drawback that legitimate traffic packets may also be dropped enroute to the destination. In [54], Kalantari et al. have proposed a proactive method for mitigation of the effects of DDoS attacks wherein each router maintains a partition of active TCP flows into aggregates. Each aggregate is probed to estimate the proportion of attack traffic that it contains. Packets belonging to aggregates that contain significant amounts of attack traffic may be subject to aggressive drop policies to prevent attack at the intended victim. Again, in this case too, legitimate packets face the risk of being dropped. For the purpose of our work, we define aggregates as a vital element of the approach. Additionally, aggregates are defined in advance of the attack so that their response measurements are taken to normal (non-attack) traffic in order to be compared later on with measurements under an attack, if any.

On the whole, studies have indicated that part of the mitigation techniques in practice today, suffer from the following drawbacks:

1. They are reactive in nature.
 2. They deploy packet dropping policies at the routers wherein even legitimate packets face the risk of being dropped.
 3. The topology of the network needs to be known in advance.
- The mitigation technique employed in the framework proposed in this paper seeks to do away with all these drawbacks as we shall see in next section.

IV. PROPOSED SOLUTION

In this section, we shall outline the various facets of our proposed framework for defense against DDoS attacks. The proposed framework provides for proactive mitigation against the effect of DDoS attacks as described next. Whenever a packet arrives at a router to be forwarded to the server to be protected from a DDoS attack, instead of sending that packet on the outbound link, a copy of its header [55] is sent toward the server for characterization. This provides a *proactive* approach to mitigation against the attack as the bandwidth of the links involved will not be exhausted by the voluminous attack traffic as only the headers (that are small in size) will traverse on the links to the server.

The technique to be used in this framework for mitigation provides the *dual* functionality of IP Traceback as well. The 16-bit IP Identification field in the header of the original packet which was being used traditionally for traceback need not be used now. In the proposed technique, the IP Identification field of the original packet will not be used for traceback purposes. Instead, the IP Identification field in the *copy of the header* generated will be used to store the edge information. The copies of headers generated represent the actual dynamics of the traffic flow to which they belong. These headers will be subject to the characterization test described next.

For classification, instead of the Welch's periodogram method used in [52], the Exactly Periodic Subspace Decomposition (EPSD) [56] technique will be used as part of this framework. The EPSD technique does away with the disadvantages of the Welch's method by difference in the selection of time domain input elements that constitute the frequency domain output elements. To get a better understanding of the proposed model, consider a sample topology shown in Figure 5. The topology considered is similar to the one used traditionally to depict a typical client-server scenario in the Internet for simulation purposes [6].

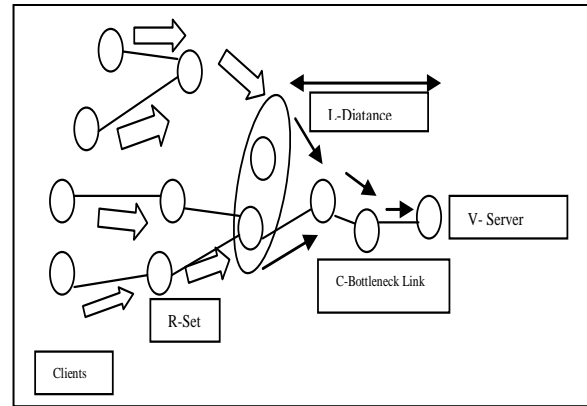


Figure 4 Proposed Framework

The clients (attack and legitimate) send their requests to the server V (indicated by thick arrows). The routers (set R) en route from the clients to the server will proactively generate copies of these packets and save the original packets with them. These routers will also stamp their identity in the Identification field of the copy of the IP header thus generated and send them to V (indicated by thin arrows). The other routers through which these header copies will traverse before reaching V will also append their edge information in the same Identification field. Once these header copies reach the bottleneck link C, they will undergo the EPSD test for periodicity and thus the flows will be characterized as attack or legitimate. If a flow is characterized as a legitimate flow, only then will the routers belonging to set R be instructed to forward the stored packets to the server. If a flow is characterized as an attack flow, then the encoding information in the generated copies of the headers will be used to construct the attack graph for IP Traceback [57] and the routers (set R) will be asked to drop the corresponding original attack packets. A flowchart depicting the solution is illustrated below.

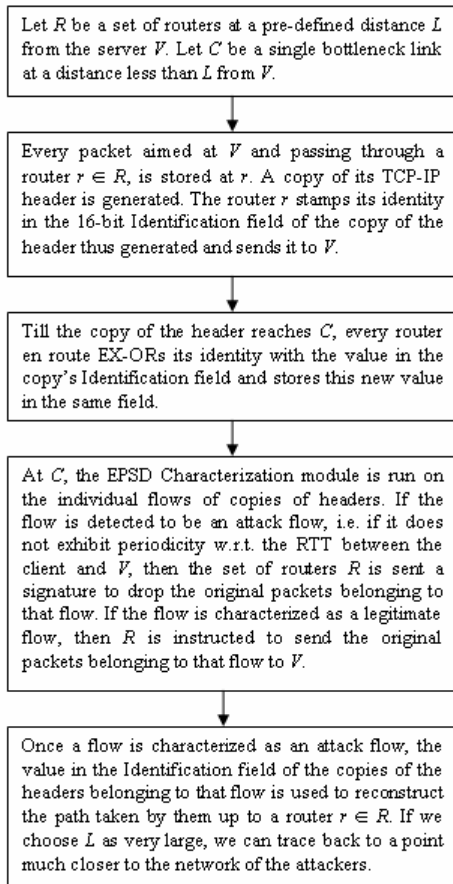


Figure 5. Flowchart of the Proposed Framework

V. CONCLUSION AND FUTURE WORK

As malicious entities unleash an increasing number of DDoS attacks on the Internet, it has become imperative to not only track them to hold them liable (traceback), but also to limit their capabilities and render them ineffective (mitigation). In this paper, we propose a novel framework that provides both traceback and mitigation capabilities. On the whole, studies have indicated that part of the mitigation techniques in practice today, suffer from the following drawbacks: they are reactive in nature; they deploy packet dropping policies at the routers wherein even legitimate packets face the risk of being dropped; the topology of the network needs to be known in advance. The mitigation technique employed in the framework proposed in this study seeks to do away with all these drawbacks.. We hope to evaluate the proposed framework and techniques on an NS2 network simulation platform in our subsequent works.

REFERENCES

- [1] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, "Reducing unwanted traffic in a backbone network," in *Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI)*, 2005, pp. 9-15.
- [2] CERT Coordination Center, "Denial of service attacks." Available at http://www.cert.org/tech_tips/denial_of_service.html, March 2007.
- [3] L. Garber, "Denial-of-service attacks rip the Internet." *IEEE Computer*, vol. 33, no. 4, April 2000, pp.12-17.
- [4] D. Moore, G. M. Voelker, and S. Savage, "Inferring Internet denial-of-service activity." in *Proceedings of the 10th USENIX Security Symposium*, August 2001.
- [5] CAIDA, "Nameserver DoS attack October 2002." Available at <http://www.caida.org/funding/dns-analysis/oct02dos.xml>, March 2007.
- [6] ZDNet, "New denial-of-service threat emerges." Available at <http://news.zdnet.com/2100-1009-22-6050688.html>, March 2007.
- [7] CERT. Denial of Service Attacks. http://www.cert.org/tech_tips/denial_of_service.html, 1997.
- [8] K.J. Houle, G. M. Weaver, N. Long, and R. Thomas. "Trends in denial of service attack technology". *Technical Report Version 1.0*, CERT Coordination Center, Carnegie Mellon University (2001) http://www.cert.org/archive/pdf/DoS_trends.pdf.
- [9] F. Lau, S. H. Rubin, M. H. Smith, and L. Trajkovi. "Distributed denial of service Attacks," in *Proceedings of IEEE International Conference on Systems and Cybernetics*, vol. 3, pp. 2275-2280, 2000.
- [10] J. MäolsÄa, "Mitigating denial of service attacks in computer networks". PhD thesis, Helsinki University of Technology, Espoo, Finland, June 2006.
- [11] V. Paxson, "An analysis of using reflectors for distributed denial-of-service attacks." *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 3, July 2001.
- [12] S. M. Specht and R. B. Lee, "Distributed denial of service: taxonomies of attacks, tools and countermeasures." in *Proceedings of the 17th International Conference on Parallel and Distributed Computing Systems*, September 2004, pp. 543-550.
- [13] R. K. Chang, "Defending against flooding-based distributed denial-of-service attacks: A tutorial." *IEEE Commun. Mag.*, vol. 40, no. 10, October 2002, pp.42-51.
- [14] Cisco Systems, Inc., "Characterizing and tracing packet floods using cisco routers." May 2005. Attacks", in *IEEE TPDS*, 2003.
- [15] H. Burch, B. Cheswick, "Tracing Anonymous Packets to their approximate source", in *Proc. USENIX LISA*, Dec. 2000.
- [16] S. M. Bellovin, "ICMP traceback messages," Internet Draft, Mar. 2000.
- [17] A. C. Snoeren et. al., "Hash-Based IP Traceback," in *SIGCOMM*, 2001.
- [18] Li et. al., "Large-Scale IP Traceback in High-Speed Internet: Practical Techniques & Theoretical Foundation", in *IEEE Symp. on Security & Privacy* 2004.
- [19] A. Belenky, N. Ansari, "IP Traceback with Deterministic Packet Marking", in *IEEE Communication Letters*, Apr 2003.
- [20] Savage et. al., "Practical Network Support for Traceback," in *SIGCOMM*, 2000.
- [21] D. Song, A. Perrig, "Advanced and Authenticated Marking Schemes for IP traceback", in *Proc. IEEE INFOCOM*, 2001.
- [22] M. Adler, "Tradeoffs in Probabilistic packet marking for IP traceback", in *Proc. STOC*, pp.407-418, 2002.
- [23] T. Peng, C. Leckie, K. Ramamohanarao, "Adjusted Probabilistic Packet Marking for IP Trace-back", in *Proc. Networking*, 2002.
- [24] A. Yaar, A. Perrig, D. Song, "FIT: Fast Internet Traceback", in *INFOCOM* 2005.
- [25] I. Hamadeh, G. Kesidis, "Performance of IP Address Fragmentation Strategies for DDoS trace-back", in *Proc. IEEE IPOM*, 2003.

- [26] M. Waldvogel, "GOSSIB vs Traceback Rumors", in ACSAC, 2002.
- [27] M. Ma, "Tabu Marking Scheme for Traceback", in IPDPS, 2005.
- [28] B. Al-Duwairi, T. Daniels, "Topology Based Packet Marking," in ICCCN 2004.
- [29] M. Muthuprasanna, G. Manimaran, "Space-Time Encoding for DDoS Attack Traceback", in Proc. IEEE GLOBECOM, 2005.
- [30] M. Muthuprasanna, G. Manimaran, Mansoor Alicherry, Vijay Kumar, "Coloring the Internet: IP Traceback", in Proc. ICPADS, 2006.
- [31] K. Choi, H. Dai, "A Marking Scheme using Huffman Codes for IP Traceback", in Proc. ISPAN, 2004.
- [32] C. Bai et.al., "Algebraic Geometric Code Based IP Traceback", in IPCCC 2004.
- [33] Y. Sawai, M. Oe, K. Iida, Y. Kadobayashi, "Performance Evaluation of Inter-Domain IP Traceback", in Proc. IEEE ICT, 2003.
- [34] B. Al-Duwairi, G. Manimaran, "Novel Hybrid Schemes employing Packet Marking & Logging for Traceback", in IEEE TPDS, 2005.
- [35] Gong et.al., "IP Traceback based on Packet Marking & Logging", in ICC 2005.
- [36] M. Walfish, M. Vutukuru, Hari Balakrishnan, D. Karger, Scott Shenker, "DDoS Defense by Offense", in Proc. SIGCOMM, 2006.
- [37] P. Ferguson, D. Senie, "Network ingress filtering: Defeating denial of service attacks which employ IP source address spoofing," in RFC 2267, 1998.
- [38] J. Li, J. Mirkovic, M. Wang, M. Reiher, L. Zhang, "SAVE: Source address validity enforcement protocol", in Proc. of INFOCOM, 2001.
- [39] R. Stone, "CenterTrack: An IP overlay network for tracking DoS floods", in Proc. USENIX Security Symposium, 2000.
- [40] M. Sung, J. Xu, "Intelligent Packet Filtering: A Novel Technique for defending against DDoS Attacks", in IEEE TPDS, 2003.
- [41] A. Yaar, A. Perrig, D. Song, "Pi: A Path Identification Mechanism to defend against DDoS Attacks," in Proc. IEEE Symposium on Security and Privacy, 2003.
- [42] A. Yaar, A. Perrig, D. Song, "StackPi: New Packet Marking Filtering Mechanisms for DDoS & IP Spoofing Defense", in JSAC, pp.1853-1863, Oct. 2006.
- [43] C. Jin, H. Wang, K. G. Shin, "Hop-Count Filtering: An effective defense against spoofed DDoS traffic", in ACM CCS, 2003.
- [44] A. Keromytis, V. Misra, D. Rubenstein, "SOS: An architecture for mitigating DDoS attacks", in IEEE JSAC, pp. 176-188, Jan. 2004.
- [45] Y. Kim, W. Lau, M. Chuah, J. Chao, "PacketScore: A statistical-based overload control against DDoS attacks", in Proc. IEEE INFOCOM, 2004.
- [46] T. Peng, C. Leckie, K. Ramamohanarao, "Protection from DDoS attacks using history-based IP filtering", in Proc. IEEE ICC, 2003.
- [47] J. Ioannidis, S. M. Bellovin, "Implementing Pushback: Router-based defense against DDoS attacks", in Proc. NDSS, 2002.
- [48] D. Yau, J. Lui, F. Liang, "Defending against DDoS attacks with max-min fair server-centric router throttles", in Proc. IWQoS, 2002.
- [49] Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson, "Practical network support for ip traceback," in *Proceedings of the 2000 ACM SIGCOMM Conference*, Aug. 2000.
- [50] Dawn Xiaodong Song and Perrig A., "Advanced and authenticated marking schemes for IP traceback," *Proceedings. IEEE INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 2, pp: 878 – 886, Apr. 2001.
- [51] H. Wang, D. Zhang, and K. G. Shin. "Detecting SYN flooding attacks," *In Proceedings of IEEE INFOCOM 2002*, pp. 1530 – 1539, June 2002.
- [52] Chen-Mou Cheng, H. T. Kung, and Koan-Sin Tan, "Use of Spectral Analysis in Defense Against DoS Attacks," *In the Proceedings of Global Telecommunications Conference, 2002, GLOBECOM '02. IEEE*, Vol. 3, pp: 2143 – 2148, Nov. 2002.
- [53] B. Stephan, "Optimal filtering for denial of service mitigation," *Proceedings of the 41st IEEE Conference on Decision and Control*, 2002, Vol. 2, pp: 1428 – 1433, Dec. 2002.
- [54] M. Kalantari, K. Gallicchio and M. A. Shayman, "Using transient behavior of TCP in mitigation of distributed denial of service attacks," *Proceedings of the 41st IEEE Conference on Decision and Control*, 2002, Vol. 2, pp: 1422 – 1427, Dec. 2002.
- [55] S. Bellovin. ICMP traceback messages, March 2000. Internet Draft: <http://www.cs.columbia.edu/~smb/papers/draft-bellovinitrace-00.txt>
- [56] D. Darian Muresan, and Thomas W. Parks, "Orthogonal, Exactly Periodic Subspace Decomposition," *IEEE Trans. On Signal Processing*, Vol. 51, No. 9, pp: 2270-2279, Sep. 2003.
- [57] Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson, "Practical network support for ip traceback," in *Proceedings of the 2000 ACM SIGCOMM Conference*, Aug. 2000.

Accurate and efficient crawling

The deep web: Surfacing hidden value

SUNEET KUMAR

Associate Professor ;Computer Science Dept.
Dehradun Institute of Technology,
Dehradun,India
suneetcit81@gmail.com

ANUJ KUMAR YADAV

Assistant Professor ;Computer Science Dept.
Dehradun Institute of Technology,
Dehradun,India
anujbit@gmail.com

RAKESH BHARATI

Assistant Professor ;Computer Science Dept.
Dehradun Institute of Technology,
Dehradun,India
goswami.rakesh@gmail.com

RANI CHOUDHARY

Sr. Lecturer;Computer Science Dept.
Babu Bnarsi Das Institute of Technology,
Ghaziabad,India
ranichoudhary04@gmail.com

Abstract—Searching Focused web crawlers have recently emerged as an alternative to the well-established web search engines. While the well-known focused crawlers retrieve relevant web-pages, there are various applications which target whole websites instead of single web-pages. For example, companies are represented by websites, not by individual web-pages. To answer queries targeted at websites, web directories are an established solution. In this paper, we introduce a novel focused website crawler to employ the paradigm of focused crawling for the search of relevant websites. The proposed crawler is based on two-level architecture and corresponding crawl strategies with an explicit concept of websites. The external crawler views the web as a graph of linked websites, selects the websites to be examined next and invokes internal crawlers. Each internal crawler views the web-pages of a single given website and performs focused (page) crawling within that website. Our experimental evaluation demonstrates that the proposed focused website crawler clearly outperforms previous methods of focused crawling which were adapted to retrieve websites instead of single web-pages.

Keywords—Deep Web;Quality Documents; Surface Web;Topical Database .

I. INTRODUCTION

A. The Deep Web

Internet content is considerably more diverse and the volume certainly much larger than commonly understood. First, though sometimes used synonymously, the World Wide Web is but a subset of Internet content. Other Internet protocols besides the Web include FTP (file transfer protocol), e-mail, news, Telnet, and Gopher (most prominent among pre-Web protocols). This paper does not consider further these non-Web protocols [1]. Second, even within the strict context of the Web, most users are aware only of the content presented to them via search engines such as Excite, Google, AltaVista, or

Northern Light, or search directories such as Yahoo!, About.com, or Look Smart. Eighty-five percent of Web users use search engines to find needed information, but nearly as high a percentage cite the inability to find desired information as one of their biggest frustrations.[2] According to a recent survey of search-engine satisfaction by market-researcher NPD, search failure rates have increased steadily since 1997.[3] The importance of information gathering on the Web and the central and unquestioned role of search engines -- plus the frustrations expressed by users about the adequacy of these engines -- make them an obvious focus of investigation.

II. GENERAL DEEP WEB CHARACTERISTICS

Deep Web content has some significant differences from surface Web content. Deep Web documents (13.7 KB mean size; 19.7 KB median size) are on average 27% smaller than surface Web documents. Though individual deep Web sites have tremendous diversity in their number of records, ranging from tens or hundreds to hundreds of millions (a mean of 5.43 million records per site but with a median of only 4,950 records), these sites are on average much, much larger than surface sites. The rest of this paper will serve to amplify these findings. This mean deep Web site has a Web-expressed (HTML-included basis) database size of 74.4 MB (median of 169 KB). Actual record counts and size estimates can be derived from one-in-seven deep Web sites. On average, deep Web sites receive about half again as much monthly traffic as surface. The median deep Web site receives somewhat more than two times the traffic of a random surface Web site (843,000 monthly page views vs. 365,000). Deep Web sites on average are more highly linked to than surface sites by nearly a factor of two (6,200 links vs. 3,700 links), though the median deep Web site is less so (66 vs. 83 links). This

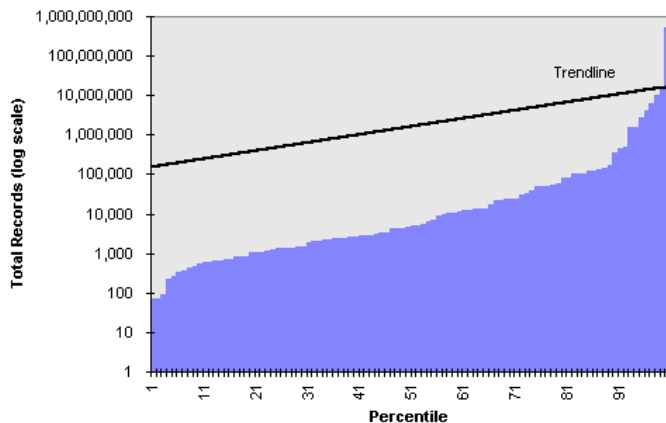
suggests that well-known deep Web sites are highly popular, but that the typical deep Web site is not well known to the Internet search public. One of the more counter-intuitive results is that 97.4% of deep Web sites are publicly available without restriction; a further 1.6% is mixed (limited results publicly available with greater results requiring subscription and/or paid fees); only 1.1% of results are totally subscription or fee limited. This result is counter intuitive because of the visible prominence of subscriber-limited sites such as Dialog, Lexis-Nexis, Wall Street Journal Interactive, etc. (We got the document counts from the sites themselves or from other published sources.) However, once the broader pool of deep Web sites is looked at beyond the large, visible, fee-based ones, public availability dominates.

Deep Web sites contain data of about 750 terabytes (HTML-included basis) or roughly forty times the size of the known surface Web. These sites appear in a broad array of domains from science to law to images and commerce. We estimate the total number of records or documents within this group to be about eighty-five billion. Roughly two-thirds of these sites are public ones, representing about 90% of the content available within this group of sixty. The absolutely massive size of the largest sites shown also illustrates the universal power function distribution of sites within the deep Web, not dissimilar to Web site popularity [3] or surface Web sites. One implication of this type of distribution is that there is no real upper size boundary to which sites may grow.

III. DISTRIBUTION OF DEEP WEB SITES

The under count due to lack of randomness and what we believe to be the best estimate above, namely the Lycos-Info mine pair, indicate to us that the ultimate number of deep Web sites today is on the order of 200,000.

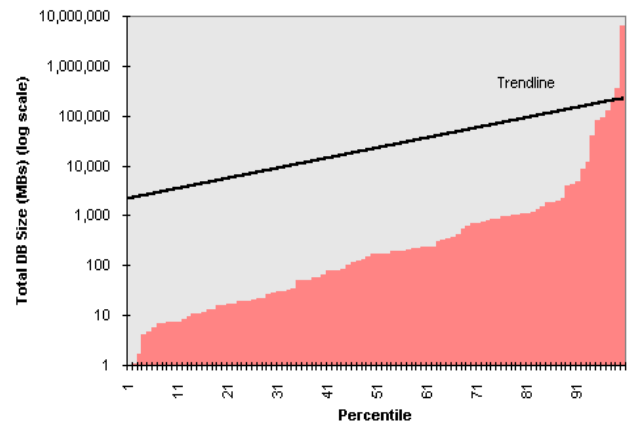
Figure 1. Inferred Distribution of Deep Web Sites, Total Record Size



Plotting the fully characterized random 100 deep Web sites against total record counts produces Figure 1. Plotting these same sites against database size (HTML-included basis) produces Figure 2. Multiplying the mean size of 74.4 MB per deep Web site times a total of 200,000 deep Web sites results in a total deep Web size projection of 7.44 pet bytes, or 7,440

terabytes. [5][6a] Compared to the current surface Web content estimate of 18.7 TB, this suggests a deep Web size about 400 times larger than the surface Web., the deep Web size calculates as 120 times larger than the surface Web. At the highest end of the estimates, the deep Web is about 620 times the size of the surface Web. Alternately, multiplying the mean document/record count per deep Web site of 5.43 million times 200,000 total deep Web sites results in a total record count across the deep Web of 543 billion documents.[6b] Compared to the estimate of one billion documents, this implies a deep Web 550 times larger than the surface Web. At the low end of the deep Web size estimate this factor is 170 times; at the high end, 840 times. Clearly, the scale of the deep Web is massive, though uncertain. Since 60 deep Web sites alone are nearly 40 times the size of the entire surface Web, we believe that the 200,000 deep Web site bases is the most reasonable one. Thus, across database and record sizes, we estimate the deep Web to be about 500 times the size of the surface Web.

Figure 2. Inferred Distribution of Deep Web Sites, Total Database Size (MBs)



IV. DEEP WEB COVERAGE IS BROAD, RELEVANT

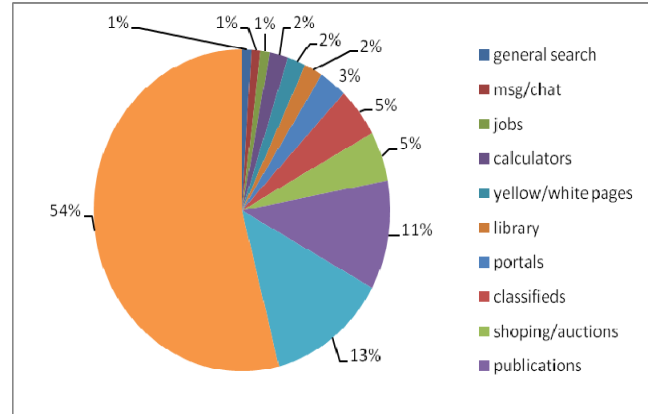
The subject coverage across all 17,000 deep Web sites used in this study. These subject areas correspond to the top-level subject structure of the Complete-Planet site. The table shows a surprisingly uniform distribution of content across all areas, with no category lacking significant representation of content. Actual inspection of the Complete-Planet site by node shows some subjects are deeper and broader than others. However, it is clear that deep Web content also has relevance to every information need and market.

Table 1. Distribution of Deep Sites by Subject Area

<i>Deep web coverage</i>	
<i>Agriculture</i>	2.7%
<i>Arts</i>	6.6%
<i>Business</i>	5.9%
<i>Computing Web</i>	6.9%
<i>Education</i>	4.3%
<i>Employment</i>	4.1%
<i>Engineering</i>	3.1%
<i>Government</i>	3.9%
<i>Health</i>	5.5%
<i>Humanities</i>	13.5%
<i>Law/polices</i>	3.9%
<i>Lifestyle</i>	4.0%
<i>News/Media</i>	12.2%
<i>People, companies</i>	4.9%
<i>Recreation, Sports</i>	3.5%
<i>References</i>	4.5%
<i>Science, Math</i>	4.0%
<i>Travel</i>	3.4%
<i>Shopping</i>	3.2%

More than half of all deep Web sites feature topical databases. Topical databases plus large internal site documents and archived publications make up nearly 80% of all deep Web sites. Purchase-transaction sites - including true shopping sites with auctions and classifieds -- account for another 10% or so of sites. The other eight categories collectively account for the remaining 10% or so of sites.

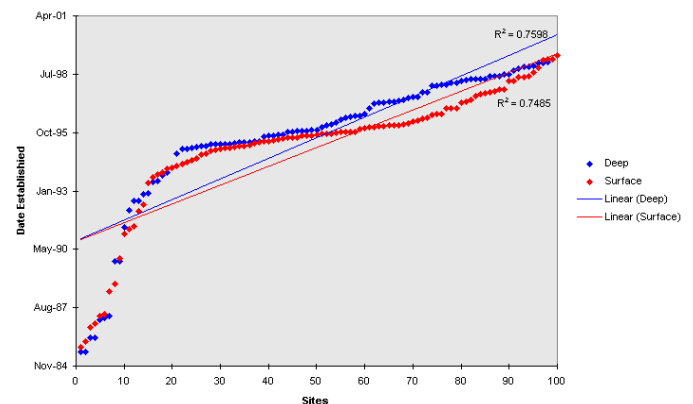
Figure 3. Distribution of Deep Web Sites by Content Type



A. Deep Web Growing Faster than Surface Web

Lacking time-series analysis, we used the proxy of domain registration date to measure the growth rates for each of 100 randomly chosen deep and surface Web sites. These results are presented as a scatter gram with superimposed growth trend lines in Figure 4

Figure 4. Comparative Deep and Surface Web Site Growth Rates



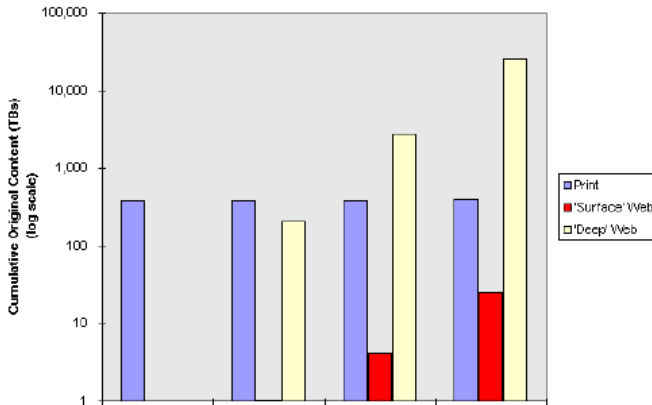
. Use of site domain registration as a proxy for growth has a number of limitations. First, sites are frequently registered well in advance of going "live." Second, the domain registration is at the root or domain level (e.g., www.mainsite.com). The search function and page -- whether for surface or deep sites -- often is introduced after the site is initially unveiled and may itself reside on a subsidiary form not discoverable by the who is analysis.

V. ORIGINAL DEEP CONTENT NOW EXCEEDS ALL PRINTED GLOBAL CONTENT

International Data Corporation predicts that the number of surface Web documents will grow from the current two billion or so to 13 billion within three years, a factor increase of 6.5 times;[7] deep Web growth should exceed this rate, perhaps increasing about nine-fold over the same period. Figure 4

compares this growth with trends in the cumulative global content of print information drawn from a recent UC Berkeley study.[8a]

Figure 5. 15-yr Growth Trends in Cumulative Original Information Content (log scale)



The total volume of printed works (books, journals, newspapers, newsletters, office documents) has held steady at about 390 terabytes (TBs).[8b] By about 2000, deep Web original information content equaled all print content produced through history up until that time. By 2005, original deep Web content is estimated to have exceeded print by a factor of seven and is projected to exceed print content by a factor of sixty three by 2011. Other indicators point to the deep Web as the fastest growing component of the Web and will continue to dominate it. [9] Even today, at least 240 major libraries have their catalogs on line; UMI, a former subsidiary of Bell & Howell, has plans to put more than 5.5 billion document images online;[11] and major astronomy data initiatives are moving toward putting pet bytes of data online. [12] These trends are being fueled by the phenomenal growth and cost reductions in digital, magnetic storage.[8c][13] International Data Corporation estimates that the amount of disk storage capacity sold annually grew from 10,000 terabytes in 1994 to 116,000 terabytes in 2000, and it is expected to increase to 1,400,000 terabytes in 2011.[14] Deep Web content accounted for about 1/338th of magnetic storage devoted to original content in 2010; it is projected to increase to 1/200th by 2011. As the Internet is expected to continue as the universal medium for publishing and disseminating information, these trends are sure to continue.

VI. DEEP VS. SURFACE WEB QUALITY

The issue of quality has been raised throughout this study. A quality search result is not a long list of hits, but the right list. Searchers want answers. Providing those answers has always been a problem for the surface Web, and without appropriate technology will be a problem for the deep Web as well. Effective searches should both identify the relevant information desired and present it in order of potential relevance -- quality. Sometimes what is most important is comprehensive discovery -- everything referring to a commercial product, for instance. Other times the most

authoritative result is needed -- the complete description of a chemical compound, as an example. The searches may be the same for the two sets of requirements, but the answers will have to be different. Meeting those requirements is daunting, and knowing that the deep Web exists only complicates the solution because it often contains useful information for either kind of search. If useful information is obtainable but excluded from a search, the requirements of either user cannot be met. We have attempted to bring together some of the metrics included in this paper,[15] defining quality as both actual quality of the search results and the ability to cover the subject.

Table 2. Total "Quality" Potential, Deep vs. Surface Web

Search Type	Total Docs(Million)	Quality Docs(Million)
Surface Web		
Single Site search	160	7
Meta-site search	840	38
Total Surface possible	1,000	45
Deep web		
Mega deep Search	110,000	14,850
Total deep Possible	550,000	74,250
Deep v. Surface web Improving Ratio		
Single Site Search	668:1	2,063:1
Meta-site search	131:1	393:1
Total Possible	655:1	2,094:1

Web sites may improve discovery by 600 fold or more. Surface Web sites are fraught with quality problems. For example, a study in 1999 indicated that 44% of 1998 Web sites were no longer available in 1999 and that 45% of existing sites were half-finished, meaningless, or trivial.[16] Lawrence and Giles' NEC studies suggest that individual major search engine coverage dropped from a maximum of 32% in 1998 to 16% in 1999.[4] Peer-reviewed journals and services such as Science Citation Index have evolved to provide the authority necessary for users to judge the quality of information. The Internet lacks such authority. An intriguing possibility with the deep Web is that individual sites can themselves establish that authority. For example, an archived publication listing from a peer-reviewed journal such as Nature or Science or user-accepted sources such as the Wall Street Journal or The Economist carry with them authority based on their editorial and content efforts. The owner of the site vets what content is made available. Professional content suppliers typically have the kinds of database-based sites that make up the deep Web; the static HTML pages that typically make up the surface Web are less likely to be from professional content suppliers. By directing queries to deep Web sources, users can choose authoritative sites. Search engines, because of their indiscriminate harvesting, do not direct queries. By careful selection of searchable sites, users can make their own determinations about quality, even though a solid metric for that value is difficult or impossible to assign universally.

VII. CONCLUSION

Serious information seekers can no longer avoid the importance or quality of deep Web information. But deep Web information is only a component of total information available. Searching must evolve to encompass the complete Web. Directed query technology is the only means to integrate deep and surface Web information. The information retrieval answer has to involve both "mega" searching of appropriate deep Web sites and "meta" searching of surface Web search engines to overcome their coverage problem. Client-side tools are not universally acceptable because of the need to download the tool and issue effective queries to it.[17] Pre-assembled storehouses for selected content are also possible, but will not be satisfactory for all information requests and needs. Specific vertical market services are already evolving to partially address these challenges.[18] These will likely need to be supplemented with a persistent query system customizable by the user that would set the queries, search sites, filters, and schedules for repeated queries. These observations suggest a splitting within the Internet information search market: search directories that offer hand-picked information chosen from the surface Web to meet popular search needs; search engines for more robust surface-level searches; and server-side content-aggregation vertical "info-hubs" for deep Web information to provide answers where comprehensiveness and quality are imperative.

REFERENCES

- [1]. A couple of good starting references on various Internet protocols can be found at <http://wdvl.com/Internet/Protocols/> and http://www.webopedia.com/Internet_and_Online_Services/Internet/Internet_Protocols/.
- [2]. Tenth edition of Gvu's (graphics, visualization and usability) WWW User Survey, May 14, 1999. [formerly http://www.gvu.gatech.edu/user_surveys/survey-1998-10/tenthreport.html.]
- [3]. "4th Q NPD Search and Portal Site Study," as reported by Search Engine Watch [formerly <http://searchenginewatch.com/reports/npd.html>]. NPD's Web site is at <http://www.npd.com/>.
- [4]. S. Lawrence and C.L. Giles, "Accessibility of Information on the Web," *Nature* 400:107-109, July 8, 1999.
- [5]. 1024 bytes = 1 kilobyte (KB); 1000 KB = 1 megabyte (MB); 1000 MB = 1 gigabyte (GB); 1000 GB = 1 terabyte (TB); 1000 TB = 1 pet byte (PB). In other words, 1 PB = 1,024,000,000,000,000 bytes or 10^{15} .
- [6]. 6a, 6b. Our original paper published on July 26, 2000, used estimates of one billion surface Web documents and about 100,000 deep Web sea reachable databases. Since publication, new information suggests a total of about 200,000 deep Web searchable databases. Since surface Web document growth is now on the order of 2 billion documents, the ratios of surface to Web documents (400 to 550 times greater in the deep Web) still approximately holds. These trends would also suggest roughly double the amount of deep Web data storage to fifteen pet bytes than is indicated in the main body of the report.
- [7]. As reported in Sequoia Software's IPO filing to the SEC, March 23, 2000; see <http://www.10kwizard.com/filing.php?repo=tenk&image=1117423&doc=1&total=266&back=2&g=>.
- [8]. 8a, 8b, 8c. P. Lyman and H.R. Varian, "How Much Information," published by the UC Berkeley School of Information Management and Systems, October 18, 2000. See <http://www.sims.berkeley.edu/research/projects/how-much-info/index.html>. The comparisons here are limited to achievable and retrievable public information, exclusive of entertainment and communications content such as chat or e-mail.
- [9]. As this analysis has shown, in numerical terms the deep Web already dominates. However, from a general user perspective, it is unknown.
- [10]. See <http://lcweb.loc.gov/z3950/>.
- [11]. See <http://www.infotoday.com/newsbreaks/nb0713-3.htm>.
- [12]. A. Hall, "Drowning in Data," *Scientific American*, Oct. 1999; [formerly <http://www.sciam.com/explorations/1999/100499data/>].
- [13]. As reported in Sequoia Software's IPO filing to the SEC, March 23, 2000; see <http://www.10kwizard.com/filing.php?repo=tenk&image=1117423&doc=1&total=266&back=2&g=>.
- [14]. From Advanced Digital Information Corp., Sept. 1, 1999, SEC filing; [formerly ["\]>\[http://www.tenkwizard.com/fil_blurb.asp?iacc=991114&exp=terabytes%20and%20online&g=\]\(http://www.tenkwizard.com/fil_blurb.asp?iacc=991114&exp=terabytes%20and%20online&g=\)\].](http://www.tenkwizard.com/fil_blurb.asp?iacc=991114&exp=terabytes%20and%20online&g=)
- [15]. Assumptions: SURFACE WEB: for single surface site searches - 16% coverage; for meta search surface searchers - 84% coverage [higher than NEC estimates in reference 4; based on empirical Bright-Planet searches relevant to specific topics]; 4.5% quality retrieval from all surface searches. DEEP WEB: 20% of potential deep Web sites in initial Complete-Planet release; 200,000 potential deep Web sources; 13.5% quality retrieval from all deep Web searches.
- [16]. Online Computer Library Center, Inc., "June 1999 Web Statistics," Web Characterization Project, OCLC, July 1999. See the Statistics section in <http://wcp.oclc.org/>.
- [17]. Most surveys suggest the majority of users are not familiar or comfortable with Boolean constructs or queries. Also, most studies suggest users issue on average 1.5 keywords per query; even professional information scientists issue 2 or 3 keywords per search. See further Bright-Planet's search tutorial at <http://www.completeplanet.com/searchresources/tutorial.htm>.
- [18]. See, as one example among many, CareData.com, at [formerly http://www.citeline.com/pro_info.html]. L.

MOBILE PHONE AUGMENTED REALITY BUSINESS CARD

EDMUND NG GIAP WENG

Centre of Excellence for Semantic Technology and
Augmented Reality, Faculty of Cognitive Sciences and
Human Development, Universiti Malaysia Sarawak
nggiapweng@yahoo.com

TEO TZONG REN

Centre of Excellence for Semantic Technology and
Augmented Reality, Faculty of Cognitive Sciences and
Human Development, Universiti Malaysia Sarawak

BEHRANG PARHIZKAR

Faculty of Information & Communication Technology,
LIMKOKWING University of Creative technology
Cyberjaya, Selangor, Malaysia
hani.pk@limkokwing.edu.my

ARASH HABIBI LASHKARI

Faculty of Information & Communication Technology,
LIMKOKWING University of Creative technology
Cyberjaya, Selangor, Malaysia
a_habibi_1@hotmail.com

ABSTRACT- The main idea for this project is to introduce the technology of Mobile Phone Augmented Reality (AR) and develop an application for business card. AR technology is rather easier to introduce to the public by using Mobile Phone (Symbian OS). Thus, this project can bring up the basic idea of Mobile Phone AR technology and its application on a physical Business Card. Virtual 3D map is the information that display on Mobile Phone by using AR technology, while typical Business Card functions as marker which assist in marker tracking and location identifying process. Furthermore, lower-end graphic is the challenge of this project in order to robust the capability of Mobile Phone followed by enhances efficiency rendering process in real time with AR technology.

I. INTRODUCTION

Augmented Reality (AR) technology, is a growing area that originated from the Virtual Reality (VR) technology, yet vary from VR where the virtual objects is superimposed upon or composite in the real world and the users are interacting with the virtual objects in real-time (Vallino, 1998). AR require 6DOF pose tracking of device such as head-mounted displays (HMD), tangible interface object, etc; where the pose tracking must be inexpensive, work robustly and in time in changing environmental conditions (Wagner & Schmalstieg, 2009). However, the trend of AR device has switched from the first mobile augmented reality system - Backpack with HMD (Figure 1a) to a small and low cost device - Smart Phone (Figure 1d) (Rosenblum & Julier, 2009).

The term smart phone was initially coined by unknown marketing strategists to refer to a then-new class of the cell phones that could facilitate data access and processing with significant computing power (Zheng & Ni, 2006). According to Zheng and Ni (2006), a smart phone is like a small, networked computer in the form of a cell phone which usually provides personal information management (PIM) applications and some wireless communication capacity.

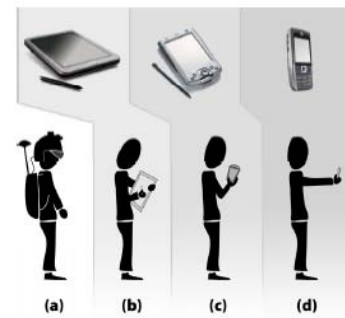


Figure 1: The evolution and miniature of mobile AR: (a) Backpack with HMD, (b) UMPC, (c) handheld, (d) Mobile phone.

Wagner and Schmalstieg (2009) stated that smart phones are aiming for a different market in AR compared to more powerful and larger ultra mobile PCs (UMPCs, Figure 1b). The smart phones are designed for a large of consumer base and mobile, yet surprisingly robust and foolproof although its appearance is fragile. Achieving sufficient performance for AR applications are therefore require careful choice of algorithms and optimized code as smart phone has limited processing capabilities compared to the PC platform (Wagner & Schmalstieg, 2009). Most of the smart phones are come with built in camera which naturally lends itself to computer vision approaches. The quality of AR in smart phone is therefore lower than AR in PC platform as the computer vision's quality is based on camera and image sensor characteristics (Wagner & Schmalstieg, 2009). Moreover, marketing has driven the development of smart phone to more megapixels rather than higher video quality (Wagner & Schmalstieg, 2009).

II. PROBLEM STATEMENT

Typical business cards come with aspect ratios of dimensions range from 1.43 to 1.8. In United State and Canada, the size of a business card is 3.5 inch x 2.0 inch or equal to 1.75 aspect ratio (PrintingForLess.com, 2009).

These sizes of business card often includes the giver's name, company affiliation (with logo) and contact information such as addresses, contact number(s), e-mail addresses and website. Today, a professional business card also includes one or more aspects of striking visual design such as map of the address location.

The standard size of a typical business card is unable to allocate more information due to the limitation of space. Sometimes, the information that included in a business card is unclear or incomplete, hence creating confusion to its viewer. An example of such is the map which is attached in a business card can sometimes confuse its viewer where the location stated is unclear.

Hence, there is a need to find a way to solve such problem. With the task at hand, this project is to design and develop a business card application by using Mobile Phone AR technology, for enhancing the functions of traditional business card. The idea of this application is to display clearer information by Mobile Phone AR technology within the limitation space in physical object of real world.

III. OBJECTIVES

- The aim is to design and develop a Mobile Phone Augmented Reality application for typical business cards.
- To design and develop 3D map using lower level graphics in Symbian OS Mobile Phone
- Implement 3D map into a physical business card by using Mobile Phone Augmented Reality technology.

IV. LITERATURE REVIEW

This chapter covers Augmented Reality (AR), Mobile Phone Augmented Reality, Business cards and Mobile 3D maps. Apart from that, some examples of existing augmented business card also will be discussed. The end of this chapter will then cover the Mobile Phone AR in the application of physical business card.

Introduction of Augmented Reality

Augmented Reality (AR) technology is a basic idea which superimposes graphics, audio and other sense enhancements over a real world environment in real-time (Vallino, 1998). According to Wagner & Schmalstieg (2009), AR require 6DOF pose tracking of device such as head-mounted displays (HMD), tangible interface object, etc; where the pose tracking must be inexpensive, work robustly and in real time in changing environmental conditions. AR technology aims to enhance the user's perception and interaction with the real world by implementing the real world with 3D virtual objects which appear to coexist in the same space as the real world.

Vallino (1998) stated that when AR technology reaches the maturity level, user shouldn't feel any conflict and discrepancies between the augmented environment and the rules that user normally senses the real world. Due to this, the goal of AR technology is therefore to create a system

where user has no idea in differentiate the real world and augmented virtual element. Hence, AR attempts to change those graphics to accommodate a user's head and eyes movements, so that the graphics always fit the perspective.

Besides, the need to know where the user is located in reference to his or her surroundings is the biggest challenge that face by developers in AR research (Bonsor, 2009). Another additional problem exist in AR is tracking the movement of users' eyes and heads. Bonsor (2009) stated that a tracking system has to recognize these movements and project the graphics related to the real-world environment the user is seeing at any moment. Although such problem exists, there are also some ways to increase the tracking accuracy. One of them is using multiple GPS signals like military. An AR system with GPS receiver is able to track user's position within an area precisely. There is a system called real-time kinematic GPS being developed, where it can achieve centimetre-level accuracy (Bonsor, 2009).

As long as researchers overcome the challenges that face them, AR will likely penetrate every corner of our lives. The AR technology offers many potential applications in various fields including maintenance and construction, military, instant information, and gaming. There are hundreds of potential applications for such a technology, gaming and entertainment being the most obvious ones (Bonsor, 2009).

Introduction of Mobile Phone Augmented Reality

In recent years, mobile phones had become an increasingly attractive platform for augmented reality technology (Rosenblum & Julier, 2009). It is predicted that the number of mobile phone will be sold in 2012 is 1.8 billion, and 800 of them are estimated to be smart phone (Rosenblum & Julier, 2009). The term Smart Phone was first introduced by unknown marketing strategists to refer to a then-new class of the cell phones which could facilitate data access and processing with significant computing power (Zheng & Ni, 2006). Zheng and Ni (2006) also stated that a smart phone is a small, networked computer in the form of a cell phone which usually provides personal information management (PIM) application and some wireless communication capacity.

As before the development of the mobile phone AR, some research had been done in mobile AR to replace the cumbersome backpack plus head-mounted display setups (Figure 1a) with ultra mobile PCs (UMPCs, Figure 1b). This evolution continues by replacement of UMPCs to PDAs (Figure 1c) and Smart Phone (Figure 1d). According to Wagner and Schmalstieg (2009), smart phone are aiming for different market in AR as compared to UMPCs. The smart phones are designed for a large of consumer base and mobile yet its AR performance surprisingly robust and foolproof although the appearance of the devices is fragile.

Thus, in order to achieving the sufficient performance for AR application, choosing algorithms carefully is needed (Rosenblum & Julier, 2009). This is because smart phone has lower computing capabilities as compared to the PC platform which is high end device. For instance, heavy use of template C++ code results in a prohibitive increase in code

size (Rosenblum & Julier, 2009). Normally, most smart phone has built-in camera, and this allow the device lends itself to computer vision approach. Nevertheless, the quality of computer tracking is highly influenced by camera and image sensor characteristics like frame size, update rate, or lens distortion (Wagner & Schmalstieg, 2009).

Symbian S60

The S60 Platform is a software platform for mobile phones that runs on Symbian OS which is an operating system (OS) designed for mobile devices and smart phones, with associated libraries, user interface, frameworks and reference implementation of common tools (Symbian Foundation, 2010). S60 is widely used by most of the smart phones in the world whereby it was first created and made as open source by Nokia and contributed to the Symbian Foundation (Symbian Foundation, 2010). According to Symbian Foundation (2010), the Symbian occupied 47% of smart phone platform market with 17.9 millions of handsets sold in fourth quarter of year 2008 and the figure is increasing every year.

The reason why S60 platform is preferred by most of the smart phones company is that it consists of a suite libraries and standard applications such as telephony, PIM tools and Helix based multimedia layers (Symbian Foundation, 2010). In addition, S60 software is a multivendor standard for smart phones that supports application development in C++, Java MIDP, Python and Adobe Flash. (Nokia, 2010).

There are four major release of S60 which are "Series 60" (2001), "Series 60 Second Edition" (2003), "Series 60 3rd Edition" (2005) and the most current one is "S60 5th Edition" (2008) (Nokia, 2009).

S60 Augmented Reality – The Map Tracking

Langlotz (2010) had created a software for Symbian phones that tracks maps, which are outfitted with regular grids of dots and tracked with 2,5 DOF. The map tracker performs six degree of freedom pose estimation from almost any map or other uniquely textured planar object. Small black dots are adding in the map to serve as reference point for tracking.

According to Langlotz (2010), the map tracker makes use of Studiersube Tracker's advanced features such as fast feature detection and poses estimation. Also, the map tracker runs on Symbian at frame rates of 15-30 fps for a complete application. Besides, the map tracker allows rotating the map arbitrarily and tilting the phone to $\sim 45^\circ$ as it treats maps planar as a full six degree of freedom (Langlotz, 2010). The map tracker works by using the regular Studierstube Tracker pipeline as a circle detector and a grid detector are adding in it.

The steps which are performed for every frame by the map tracker is declared by Langlotz (2010) as follows:

- **Thresholding**
The map tracker thresholds the image and automatically extracts dark areas.

- **Contour following**
The contour follower searches for connected regions.
- **Circle detection**
The circle detector checks all extracted contours (connected regions) using a simple ellipse fitting algorithm.
- **Grid detection**
All circles are then handed over to the grid detector that tries to reconstruct the regular grid of the dots on the map. The grid detector then extracts the patches between the dots at a resolution of 65x65 pixels. This high resolution for the patches is required in order to track maps with high frequencies (small structures such as buildings or streets at a size of a few millimeters), which would otherwise create aliasing and hence result in bad detection quality.
- **Template matching**
The template matcher uses a Gaussian downsampling to reduce the patch size to 32x32 pixels and compares it to all patches in the database.
- **Pose estimation**
Correctly detected cells are used for estimating the camera's pose relative to the map resulting in a 6DOF pose tracking.



Figure 2.1: Map Tracker by Langlotz (2010).

History of Mobile Phone Augmented Reality

According to Wagner & Schmalstieg (2003), ARToolKit has ported by them to Window CE and created the first self-contained AR application on an off-the-shelf embedded device. This port is then evolved to the ARToolKitPlus library later (Wagner & Schmalstieg, 2009).

In the year of 2004, Möhring, Lessig, & Bimber created a tracker for mobile phones that detects color-coded 3D marker shapes (Figure 2.1). Since the system did not take camera calibration or sub-pixel accuracy into account, thus its accuracy was very limited.



Figure 2.2: 3D markers by Möhring, Lessig, & Bimber (2004).

Whilst, Henrysson, Billingham and Ollila (2005) created a Symbian port of ARToolKit which evolved from the ARToolKitPlus source code and used for the AR Tennis

game (Figure 2.2) on mobile phones. The game is described as the first 2-player AR game on mobile phones.



Figure 2.3: AR-Tennis game by Henrysson, Billinghurst, & Ollila (2005).

Nearly the same time, Rohs & Gfeller (2004) created the VisualCodes systems (Figure 2.3) for Symbian smart phones. The system only provides simple tracking of 2D position on the screen, 1D rotation and a very coarse distance measure (Wagner & Schmalstieg, 2009).

While in the year of 2008, Wagner, Langlotz, & Schmalstieg created Studierstube Tracker which is a marker library supporting many different types of markers on mobile phones.



Figure 2.4: Visual Codes by Rohs & Gfeller (2004)

Introduction of Business Card

Business cards are small cards which are printed with business information about a company or individual. Business cards help one in giving great impression to stranger where it creates a professional and memorable impact every time.

A typical business card includes the giver's name, company affiliation and contact information such as addresses, contact numbers, e-mail addresses and etc (Figure 2.4). Nowadays, a professional business card includes one or more aspects of striking visual design such as map of the address location. The aspect ratios range of business cards is from 1.43 to 1.8 (PrintingForLess.com, 2009). This is because different countries have vary standard for the size of business cards (Table 1).



Figure 2.5: A typical business card

Table 1: Vary of standard size for Business Cards in several countries

Standard	Dimensions (mm)	Dimensions (in)	Aspect ratio
ISO 216, A8 sized	74 x 52	2.913 × 2.047	1.423
ISO/IEC 7810 ID-1, credit card sized	85.60 × 53.98	3.370 × 2.125	1.586
Italy, United Kingdom, France, Germany, The Netherlands, Spain, Switzerland	85 × 55	3.346 × 2.165	1.545
Australia, Denmark, New Zealand, Norway, Sweden	90 × 55	3.54 × 2.165	1.636
Japan	91 × 55	3.582 × 2.165	1.655
Hong Kong	90 × 54	3.543 × 2.125	1.667
Canada, United States	88.9 × 50.8	3.5 × 2	1.75
Czech Republic, Finland, Hungary, Israel, Kazakhstan, Poland, Russia, Slovakia, Ukraine, Bulgaria	90 × 50	3.543 × 1.968	1.8

Most Typical of Business Card Formats

Horizontal Format is the most common used of format for business cards. Most of the designers believed that it is easier to inject creativity in a business card with horizontal format (Print Centric, 2007). Another format which considered very traditional and conservative is the Vertical Format. Such format of business cards looks slightly awkward when name and address placed on it.

Another format of typical business cards is Square Format. A square format of business card is seldom used as not much space is left for writing content (Print Centric, 2007). Large Format is also one of the most typical business card formats which often used by big firm went more graphic techniques are applied into it. The advantage of Large Format is that it gives a very crisp and clean result of the image (Print Centric, 2007). Double Sided Format of business cards lets businessmen get the chance of giving more detailed contact information by using both sides of the card (Print Centric, 2007).

Furthermore, Folding Format is another type of business cards' format. This format of business cards also used by bigger firm as they use it to put more information regarding their company's achievements other than just contact details.

Current Research on Augmented Business Card

Augmented Business Card is a business card that comes with marker for AR technology tracking. In simple words, this business card is exactly a typical business card with the implementation of AR technology in it. Some research have been done in implementing AR into a typical business card. One example of the research is Toxin Labs' research where AR technology used in business card to display status of

social networks (e.g. Twitter), show personal portfolio, calling or direct contact through the application and etc (Jäger, 2009). Meanwhile, Wagner & Schmalstieg (2009) also done such research where they tracking a business card on a business card with AR technology (Figure 2.5).



Figure 2.6: A virtual avatar on a business card by using mobile-phone-based augmented reality

According to Wagner and Schmalstieg (2009), the marker is usually expected to be surrounded by black borders and containing a black and white pattern or image. However, the marker's border can be made very thin (down to zero) if the image inside the marker is darker (Wagner & Schmalstieg, 2009). This is because greyscale is converts from tracker internally.

Mobile 3D Map

3D maps are recognizable via a more realistic representation as compared to 2D maps. When the map's world is presented similar to the view of real world, one should be able to easily recognize the surroundings (Nurminen & Helin, 2005). Common 2D maps caused viewer having heavy cognitive load during the recognition process as the maps is not similar to the real world (Figure 2.6).



Figure 2.7: A 2D map, a 3D map and a photograph

Mobile 3D maps (Figure 2.5) are not a new idea since its implementation can be seen easily. The most challenging part of creating 3D maps in mobile phones is the devices have limited computational resources. Another challenging part would be lack of feasible programming interfaces since currently there are only OpenGL ES and JSR-184 has been emerged.

Due to its limited computing capabilities, lower graphic is the first consideration to be taken into account in designing 3D maps for mobile phone. This helps in speeding up the loading time of 3D graphics in mobile phone, as well as increase the performance of the application in real time.



Figure 2.8: Examples of Mobile 3D maps

V. DESIGN AND DEVELOPMENT

In this section, the architectural design of the application developed is discussed. More internal components of the application in the section of application architectural will be described.

For the internal components of the application, codes that are implemented are described. The main language for the code is Carbide.c++ since it is the core language for developing Symbian devices. Meanwhile, OpenGL ES is used to draw the scene of map and building, while ARToolKitPlus S60 is used to develop the application's augmented view.

VI. APPLICATION ARCHITECTURE

The application architecture of Mobile Phone AR Business Card has shown as above. It involves Mobile Devices and Mobile Backbone stage. The camera in Mobile Devices captures the scene of real time environment (RGB) and passes it to the Threshold for the process of gathering the environment scene in greyscale format.

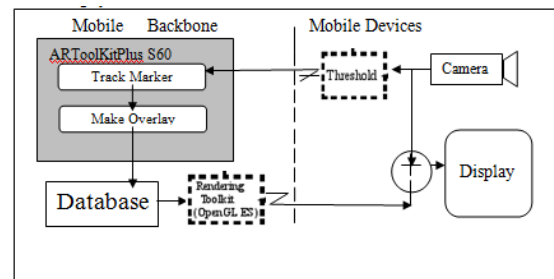


Figure 4.1: Application architecture of Mobile Phone AR Business Card

After greyscale information is gathered, ARToolKitPlus S60 is then tracks the marker (Business Card) from the scene and makes overlay. Confirming the correct location of the marker, the application is then proceed to the Database and collected information or data that are needed for drawing 3D buildings as well as map.

Moreover, the next step would be proceeding to Rendering Toolkit (OpenGL ES) in order to render the 3D map and draw them out on the marker (Business Card) in mobile phone's screen.

After all of the above processes have been through, the 3D map now draws on the Business Card by displaying on the screen of mobile phone.

Application Development

This section discusses about the development of coding and core functions which are used in the application. There were four main important parts to develop AR in Symbian based mobile. Since it is an AR based of application, video capturing was the first main part of the application development. In order to develop the AR technology in Symbian based mobile, the library of S60 3rd Edition Software Development Kit (Nokia, 2009) – ecam.lib was used for the ground part mechanism development where it operates the camera for video capturing. The second part would be the part of tracking part. ARToolKitPlus S60 (Augmented Environments Lab, 2009) was used in this part to recognize the marker then construct transform and projection matrix for rendering part. The third main part is rendering part for map and building. In order to proceed with the process of such, OpenGL ES from Khronos Group (2009) was implemented. Vertices and position of both map and building location are calculated manually and these figures were then inserted in OpenGL ES source code. In order to let the AR objects make sense, the accuracy of position between building and map is important. The fourth main part of the application development would be interaction of the AR objects and user by using keypad functions. Several functions were created by using the Carbide.c++ from Nokia (2009). These functions include the middle button (“Enter” key) works as users’ navigation and the menu list allow user to select the rain effect (with sound effect). The buttons and menu functions sources were adopted from Nokia (2009) S60 3rd Edition Software Development Kit for Symbian OS.

The application works where the map and building will be drawn by OpenGL ES and presenting on the screen once marker is detected by ARToolKitPlus S60. Here, user can use the functions that have been created to provide interaction with the application by using buttons of keypad (Carbide.c++). Generally, the application works from the combination of ARToolKitPlus S60, OpenGL ES and Carbide.c++.

Application Development’s Obstacles

There were few obstacles that existed during the development of Mobile Phone AR Business Card application. The first obstacle was limitation of OpenGL ES libraries and functions in contributing drawing part of the application. Unlike OpenGL, the extended version of it – OpenGL ES has no GLU library. Some handy functions such as gluPerspective() and glusLookAt() will have to be replaced with manual calculation. Besides, the most common immediate-mode rendering (glBegin() and glEnd()) of OpenGL also been removed in OpenGL ES and cause the widely use of vertex arrays or vertex buffers to replace it. Due to its simplified version, OpenGL ES has caused the time consuming and complexity in developing Mobile Phone AR Business Card application as huge manual calculation was essential.

Another obstacle of the application development was the application cannot be tested directly in PC by using Carbide.c++ S60 Emulator. The obstacle of such exists

because Carbide.c++ S60 Emulator does not support camera application like Mobile Phone AR Business Card. Thus, the application can only be tested by using mobile phone device for each time after application build by Carbide.c++.

VII. CODING DEVELOPMENT

This section discusses on coding development of rendering and keypad functions parts. Each of the part will be described alongside the coding are presented.

Map of UNIMAS

The first thing that needs to be implemented in the Mobile Phone AR Business Card application was the map of UNIMAS. In doing so, the map that was drawn by using GNU Image Manipulating Program (GIMP) (The GIMP Team, 2009) has saved in .jpg format and loaded in the application as texture. The image that used for UNIMAS map was named “fskpm.jpg” with the pixels size of 256*256. Stated below are the codes which been used to call and load the map image:

```
*****To load the image*****
iTextureManager = CTextureManager::NewL (iScreenWidth, iScreenHeight
FRUSTUM_TOP, FRUSTUM_BOTTOM, FRUSTUM_RIGHT,
FRUSTUM_LEFT, FRUSTUM_NEAR, this);

_LIT(Map, "fskpm.jpg");
iTextureManager->RequestToLoad(Map, &iMap );

//Start to load the textures.
iTextureManager->DoLoadL();
```

Figure 4.2: Code for texture loading

After the process of image loading, the image converted into the texture in the command of iTextureManager and the map was drawn on the screen in the drawing part. The drawing part is shown in Figure 4.2.1 where texture was set to be drawn on the screen started with certain position of vertices in the command of GLfloat vertices.

```
*****Draw Map*****
static GLubyte map[] = {0,1,2,3};
static GLfloat vertices[] = {-80.0, -55.0, 0.0,
.....
.....}

glBindTexture( GL_TEXTURE_2D, iMap.iID );
glTexParameterf( GL_TEXTURE_2D, GL_TEXTURE_MIN_FILTER,
GL_NEAREST );
glVertexPointer(3, GL_FLOAT, 0, vertices);
glTexCoordPointer(2, GL_FLOAT, 0, texCoords);
glDrawElements(GL_TRIANGLE_FAN, 4, GL_UNSIGNED_BYTE, map);
```

Figure 4.2.1: Code for texture rendering



Figure 4.2.2: Texture of UNIMAS map

The Building of Faculty of Cognitive Sciences and Human Development

The building of Faculty of Cognitive Sciences and Human Development (FCSHD) is basically developed by using the cube sample of OpenGL ES. In order to make the building looks real, the cube has to be textured with six textures which refer to six sides of the building (top, bottom, left, right, front and back). Note that every side of the cube was using the same code to draw the textures; nevertheless, their positions were set differently. The code of the building is shown as below.

```
/* Set array pointers. */
glVertexPointer( 3, GL_BYTE, 0, vertices_texture);
glTexCoordPointer( 2, GL_BYTE, 0, nokTexCoords_texture);
glNormalPointer( GL_BYTE, 0, normals_texture);

glMatrixMode( GL_TEXTURE );
glLoadIdentity();
glScalef( 1.0f/255.0f, 1.0f/255.0f, 1.0f );
glTranslatef( 128.0f, 128.0f, 0.0f );
/* Remember to change the matrix mode to modelview. */
glMatrixMode( GL_MODELVIEW );
/* Scale the cube coordinates to fit the screen. */
glScalef( 20 << 16, 20 << 16, 20 << 16 );

/* Draw the first sides of the cube with the ifskpm1.jpg texture */
glBindTexture( GL_TEXTURE_2D, ifskpm1.iID );
glTexParameterf( GL_TEXTURE_2D, GL_TEXTURE_MIN_FILTER,
GL_NEAREST );
glDrawElements( GL_TRIANGLES, 2 * 3, GL_UNSIGNED_BYTE,
triangles_top );
```

Figure 4.3: Code for FCSHD building

The building of FCSHD was added with another function where the building was set to be rotating from time to time towards y axis. The algorithms of cube rotating is shown in Figure 4.3.1.

```
/* set positions to draw cube in the map */
glTranslatef(+43.f,-27.f,-6.f);
/* Rotate the cube towards y axis */
glRotatex( aFrame << 15.0, 1 << 16.0 );
```

Figure 4.3.1: Code for FCSHD building rotating

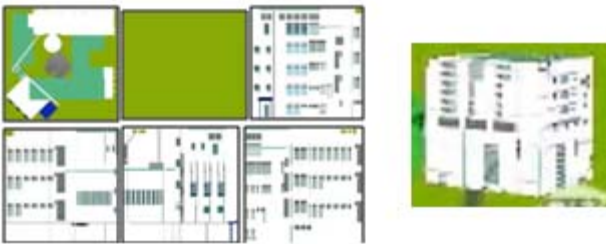


Figure 4.3.2: FCSHD textures (Top, Bottom, Front, Back, Left and Right) and the 3D building

Navigator Board

For navigator board, it is also an image that implemented in the application. The navigator board was drawn by using similar code of UNIMAS map especially for texture loading part. The function of navigation board is to inform the user about the location of FCSHD. It displays once user clicks on the “Enter” button.

```
/* Set position to draw text */
glTranslatef(64.f,-24.f,-6.f);
glBindTexture( GL_TEXTURE_2D, text.iID );
glTexParameterf( GL_TEXTURE_2D, GL_TEXTURE_MIN_FILTER,
GL_NEAREST );
glVertexPointer(3, GL_FLOAT, 0, vertices_text);
glTexCoordPointer(2, GL_FLOAT, 0, texCoords_text);
glDrawElements(GL_TRIANGLE_FAN, 4, GL_UNSIGNED_BYTE, text);
```

Figure 4.4: Code for Navigator Board



Figure 4.4.1: Texture of Navigator Board

Keypad Functions

The keypad functions were adopted from the library of S60 3rd Edition Software Development Kit under the header file of e32keys.h. It is specially design for Symbian based mobile phone. In this application, the button such as “Enter” key and “Right Select” key, “Left Select” key, “Up” key, “Down” key, “Left” key and “Right” key are used. Most of the selections were done by using Switch-case structure.

```
void CSimpleCubeAppUi::HandleCommandL(TInt aCommand)
{
    switch ( aCommand )
    {
        case EAknSoftkeyBack:
        case EEikCmdExit:
        {
            Exit();
            break;
        }
        case ESimpleCubeDecrease:
        {
            iAppContainer->iSimpleCube->VisualMode();
            break;
        }
        case ENormalMode:
        {
            iAppContainer->iSimpleCube->NormalMode();
            break;
        }
        case EAbout:
        {
            HBufC* title = iEikonEnv->AllocReadResourceLC
(R_ABOUT_DIALOG_TEXT);
            HBufC* msg = iEikonEnv->AllocReadResourceLC
(R_ABOUT_DIALOG_TITLE);
        }
        default:
        {
            break;
        }
    }
}
```

Figure 4.5: Code for Keypad Functions

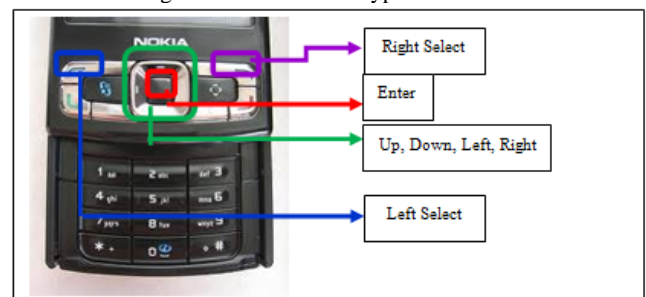


Figure 4.5.1: Keypad and buttons of Nokia N95

Rain Effect

Raining effect is an extension function that added in this application. It was implemented by adopting the source of OpenGL ES “Rain” sample. Previously, the raining effect of OpenGL ES never been used in any AR environment. The raining code that adopted from OpenGL ES was implemented by creating a Rain class for it (CRainfall). In order to render the raining effect, keypad functions were selected to control the on and off render of raining effect.

```
void CRainfall::ConstructL(GLint aParticlesCount, TVector aPosition,
                           GLfloat aWidth, GLfloat aDepth, GLfloat
                           aGroundLevel)
{
    CParticleEngine::ConstructL(aParticlesCount, aPosition);
    iWidth = aWidth;
    iDepth = aDepth;
    iGroundLevel = aGroundLevel;
    TTime now;
    now.HomeTime();
    iSeed = now.Int64();
    ResetEngine();
    iRainfall->UpdateEngine(aParticlesCount);
    iRainfall->RenderEngine(iCamera);
}
```

Figure 4.6: Code for rain effect



Figure 4.6.1: Screenshot of raining effect

Play Sound

Another extension function that included in this Mobile Phone AR Business Card application was sound effect function. In this process, the sound effect was decided to merge with the rain effect, thus raining sound clip was selected (Rain.mp3). The original OpenGL ES sample of Rain does not included with sound effect. Similar to the raining effect, the sound effect was implemented by creating a class for it (CSound) and it is controlled with the rain by keypad functions.

```
void CSound::Rain()
{
    iSound->CSound::NewL(_L("Rain.mp3"));
}
```

Figure 4.7: Code for Sound effect

VIII. DISCUSSION

The mobile phone AR application developed for the purpose of introducing the AR technology to the public easily. Thus, the basic functions that contained in PC based AR must at least applied in Mobile phone in order to proof the usefulness of AR technology.

In this Mobile Phone AR Business Card application, the basic AR flows (Video Capturing, Marker Tracking, Overlaying, and Virtual Object Rendering) were fully applied in the Symbian OS mobile phone and the application has been developed successfully.

The application works by displaying the map of UNIMAS on mobile phone screen once the marker (business card) is

detected. Then, user can interact with the application by using mobile phone keypad. The “Enter” button functions as navigator which appoints the location of FSCHD building, whilst the “Left Select” button used to display menu list which contains “3D Visuals” (raining effect), “Normal mode”, “About” and “Exit”. The building of FCSHD is rotating towards y axis in real time..



Figure 5.1: Marker (business card) detection by mobile phone



Figure 5.2: Virtual map and building displayed on mobile phone's screen



Figure 5.3: Raining effect with sound



Figure 5.4: Switch back to normal view of map by terminating the rain rendering



Figure 5.5: The “About” menu

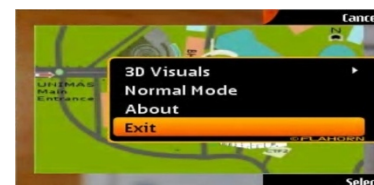


Figure 5.6: Exit or quit from application

IX. STRENGTHS OF THE APPLICATION

There are few strengths of Mobile Phone AR Business Card application. Firstly, it introduces the AR technology to the public through mobile phone. Nowadays, more and more people in the world are using mobile phone. AR technology

can easily introduce and accept by the public through mobile phone.

Secondly, the mobile phone AR solving the problem of space limitation in real world by adding in the virtual information into real world scene. Referring to the case of typical business card, it has limited spaces for locating a map. With mobile phone AR technology applied in Mobile Phone AR Business Card application, the problem solved where map is drawn on the mobile phone screen with the merge of real scene environment.

Thirdly, this application is light in size and hence, causing faster rendering in real time as compared to PC based AR technology. The reason for such is because the objects and polygons that used in mobile phone were all small in size.

The last strength for this application is it provides conveniences to the public since according to Rosenblum and Julier (2009) Symbian OS mobile phones are widely used by the public as compared to PC. Unlike mobile phone AR, PC based AR needs more set up tools such as webcam and personal PC etc. which needs more spaces and inconvenience in term of mobility.

X. WEAKNESSES OF THE APPLICATION

Low quality of graphic is one of the weaknesses in this application. Since the application was designed for mobile phone, the quality of graphics does not perform as high quality as PC due to the capability of mobile phone in handling graphics is very limited.

The complexity and limitation of OpenGL ES in drawing 3D objects is also a weakness of the application. As such, to draw a 3D object by using OpenGL ES is very complex. Hence, complex polygons cannot be drawn easily.

The next weakness is the marker for application detection in which must be an image that is surrounded with thick border lines. The reason why thick border lines are needed is because marker with thick border lines surrounded helps ARToolKitPlus S60 to focus on the interest region easily.

Screen limitation is also one of the weaknesses of the application because this limitation can easily arise since mobile phones come with small size of screen.

XI. FUTURE WORKS

Mobile Phone AR technology should not apply only in business card, but also in different fields and materials such as newspapers, leaflets, magazines and etc. By doing so, the Mobile Phone AR technology is not only exposed to public but also extends its functions in daily life. In addition, more research on implementing higher graphics in Mobile Phone AR technology are needed to maximize the performance of AR technology in Mobile Phone. Besides, more multimedia contents such as flash, video and music should also implement in this technology, thus, fully utilizes its functions and contribution to the public.

XII. CONCLUSION

In general, the application developed has achieved its objective as stated earlier in chapter (chapter 1) whereby the 3D map is implemented in Symbian OS S60 mobile phone by using the technology of Mobile Phone AR (ARToolKitPlus S60). Similar to the map tracker which introduced by Langlotz (2010), the Mobile Phone AR Business Card application able to perform well by using Nokia 95 (S60) mobile phone and presenting 3D map which is similar to the view of real world lets user easily recognize the surroundings as stated by Nurminen and Helin (2005).

The application of this project has brought the idea of implementing AR technology into a mobile phone device (Symbian OS) and represents the technology in a new way. The Mobile Phone AR Business Card works efficiently after several processes of application validation have been done to assure suitable algorithms are selected. According to Rosenblum and Julier (2009), choosing suitable algorithms carefully is needed to achieve the sufficient performance of AR application. Besides, the application also implementing AR technology in mobile phone by using low end graphics and giving a proof on the potential of mobile phone as a tool or device for introducing AR technology.

The virtual information can now be channelled everywhere by introducing Mobile Phone AR technology to the public. It does not only solve the space limitation problem in real world scene, but also augments the mobility information into real world. Other than that, mobile phone AR also provides conveniences to the public since mobile phones are widely used nowadays (Rosenblum & Julier, 2009).

ACKNOWLEDGMENT

We would like to express our appreciation to our parents and all the teachers and lecturers who help us to understand the importance of knowledge and show us the best way to gain it.

REFERENCES

- [1] ArtoolKit Marker Generator Online. (2009). MarkerGeneratorOnline.swf. Retrieved January 18, 2010, from flash.tarotaro.org.blog: <http://flash.tarotaro.org/ar/MarkerGeneratorOnline.swf>
- [2] Augmented Environments Lab. (2009). ARToolkitPlus S60 . Retrieved September 10, 2009, from Augmented Environments Lab: <http://www.augmentedenvironments.org/lab/research/tools/artoolkitplus-s60/>
- [3] Autodesk, Inc. (2009). 3ds Max 3D Animation and Rendering Software. Retrieved September 10, 2009, from Autodesk: <http://usa.autodesk.com/adsk/servlet/pc/index?id=13567410&siteID=123112>
- [4] Bonsor, K. (2009). How Augmented Reality Will Work . Retrieved September 16, 2009, from HowStuffWorks: <http://computer.howstuffworks.com/augmented-reality3.htm>

- [5] Henrysson, A., Billingham, M., & Ollila, M. (2005). Face to Face Collaborative AR on Mobile Phones. Proceedings International Symposium on Augmented and Mixed Reality (ISMAR'05), 80-89.
- [6] Jäger, J. (2009). How does it work? Retrieved September 19, 2009, from Augmented Business Card: <http://augmentedbusinesscard.net/howdoesitwork.php>
- [7] Kang, K. C., Cohen, S., Holibaugh, R., Perry, J., & Peterson, A. S. (1992). A Reuse-Based Software Development Methodology. Pittsburgh: Software Engineering Institute, Carnegie Mellon University.
- [8] Khronos Group. (2009). OpenGL ES. Retrieved September 17, 2009, from Khronos Group: <http://www.khronos.org/opengles/>
- [9] Langlotz, T. (2010). Map Tracking. Retrieved March 1, 2010, from Handheld Augmented Reality : http://studierstube.icg.tu-graz.ac.at/handheld_ar/maptracking.php
- [10] Möhring, M., Lessig, C., & Bimber, C. (2004). Video See-Through AR on Consumer Cell Phones. Proceedings of International Symposium on Augmented and Mixed Reality (ISMAR'04), 252-253.
- [11] National Physical Laboratory. (2005). Developing the Foundations for Data Management System - Software Development Cycle . Retrieved August 11, 2009, from robabdul.com: <http://www.robabdul.com/Data-Management-System-Software-Development-Cycle.asp>
- [12] Nokia. (2009). Carbide.c++. Retrieved September 18, 2009, from Forum Nokia: http://www.forum.nokia.com/Tools_Docs_and_Code/Tools/IDEs/Carbide.c++/
- [13] Nokia. (2010). Forum Nokia - Python s60. Retrieved March 1, 2010, from Nokia: http://www.forum.nokia.com/Tools_Docs_and_Code/Tools/Runtimes/Python_for_S60/
- [14] Nurminen, A., & Helin, V. (2005). TECHNICAL CHALLENGES IN MOBILE REAL-TIME 3D CITY MAPS WTH. Proceedings of IAESTED Software Engineering, Innsbruck, 1-6.
- [15] Print Centric. (2007). What are the Most Typical Business Card Formats? Retrieved September 17, 2009, from Print Centric: <http://www.printcentric.com/mosttypicalbusiness.html>
- [16] Print100. (2009, September 22). Business cards dimension standard. Retrieved September 26, 2009, from Print100.com: http://www.print100.com/hk_e/html/prod_bizcard.aspx#guide
- [17] PrintingForLess.com. (2009). Business Card Size Specifications. Retrieved September 17, 2009, from PrintingForLess.com: <http://www.printingforless.com/Business-Card-Specifications.html>
- [18] Rohs, M., & Gfeller, B. (2004). Using Camera-Equipped Mobile Phones for Interacting with Real-World Objects. Advances in Pervasive Computing, Austrian Computer Society (OCG), 265-271.
- [19] Rosenblum, L., & Julier, S. (2009, June). Projects in VR. Making Augmented Reality Practical on Mobile Phones, Part 1, 1-4.
- [20] Symbian Foundation. (2010). The Symbian Community Home. Retrieved March 03, 2010, from Symbian: <http://www.symbian.org/>
- [21] The GIMP Team. (2009). The GNU Image Manipulation Program. Retrieved August 13, 2009, from GIMP: <http://www.gimp.org/>
- [22] The Research Foundation of State University of New York . (2009). A Survey of System Development Process Models. Retrieved October 26, 2009, from Centre for Technology in Government: http://www.ctg.albany.edu/publications/reports/survey_of_sysdev?chapter=10
- [23] Vallino, J. R. (1998). Interactive Augmented Reality. Retrieved August 29, 2009, from <http://www.se.rit.edu/~jrv/publications/VallinoThesis.pdf>
- [24] Wagner, D. (2009, September 6). ARToolkitPlus. Retrieved September 16, 2009, from Handheld Augmented Reality- Christian Doppler Laboratory: http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php
- [25] Wagner, D., & Schmalstieg, D. (2007). ARToolkitPlus for Pose Tracking on Mobile Devices. Computer Vision Winter Workshop 2007, 1-8.
- [26]
- [27] Wagner, D., & Schmalstieg, D. (2003). First Steps Towards Handheld Augmented Reality. Proceedings of the 7th International Conference on Wearable Computers (ISWC 2003), 127-135.
- [28] Wagner, D., & Schmalstieg, D. (2009). History and Future of Tracking for Mobile Phone Augmented Reality. International Symposium on Ubiquitous Virtual Reality (ISUVR 2009), 1-4.
- [29] Wagner, D., & Schmalstieg, D. (2009). International Symposium on Ubiquitous Virtual Reality (ISUVR 2009). History and Future of Tracking for Mobile Phone Augmented Reality, 1-4.
- [30] Wagner, D., Langlotz, T., & Schmalstieg, D. (2008). Robust und Unobtrusive Marker Tracking on Mobile Phones. In Proc. 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, (ISMAR'08), 121-124.
- [31] Zheng, P., & Ni, L. M. (2006). Smart Phone and Next-Generation Mobile Computing. San Francisco: Morgan Kaufmann Publishers.

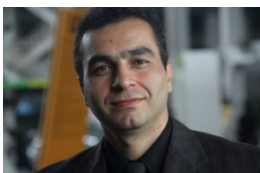


Dr. Edmund Ng Giap Weng is a lecturer at the Faculty of Cognitive Sciences and Human Development, University Malaysia Sarawak (UNIMAS). Currently, he is the Director of Centre of Excellence for Semantic Technology and Augmented Reality (CoESTAR) in UNIMAS. His expertise is in the area of Augmented Reality, Virtual Reality and Intelligent Interactive Systems. He has a research interest in the Usability Engineering:

Human Computer Interaction, Human Factors in Information Systems, and the Use of Visual Science and Artificial Intelligence to enhance trainer capability, and is overall interested in the social, organizational and technology implications of ICT deployment in organizations. He has more than 10 years experience in running multiple projects and consultations for Teaching, Research, Development and Commercialization in the universities or industries. His research team has invented ARUT Version 1.0, ARUT Mobile Version 1.0 and ARUT Web Version 1.0 for Augmented Reality Applications with international CMMI software development standard. Currently, he is involved in a number of research projects, such as Intelligent Augmented Reality application, Finger Tracking as an Input Devices; Knowledge-Based Augmented Reality for Education, Training, Museum, Planetarium, Tourism, Communication, Games and etc. He has won many national and international awards from his research projects as well as has published novel results intensively in local and international conferences, seminars, and journals in his field of interest.



Behrang Parhizkar obtained his bachelor degree in Multimedia Technology Management from Multimedia University, Malaysia, followed by master degree of Information Technology & Multimedia System, faculty of Information science and technology, University of Kebangsaan Malaysia (UKM), Bangi, Malaysia. He is a member of Visual Informatics research group. He is a principal lecturer and director of computer research center in Limkokwing University of Creative technology, Malaysia. He won some International awards such as Itex Bronz Medal of Innovation, Malaysia, Gold Medal of Invention, South Korea, and World Intellectual Property Organization (WIPO) Award for Voice Recognition Browser and Marking System. He is interested to research in Augmented Reality, Mobile Technology, Artificial Intelligent, Voice Recognition and Image Processing. He also works on Augmented Reality Technology, as a research assistant.



Arash Habibi Lashakri obtained his bachelor degree in software engineering from Azad University of

Lahijan, Gilan, Iran, followed by master degree of computer system & technology, faculty of computer science and information technology, University of Malaya (UM), Kuala Lumpur, Malaysia. Now, he is a PHD candidate of University Technology Malaysia (UTM) in Computer Science and Information Security. He is a reviewer of a few International journal of computer science and information security and Technical Committee member of some international conferences based on computer science and information Security. He is also a member of some funded research projects and he has published more than 30 papers in various international conferences and journals and three books in computer Science. He is interested to research in network security, cryptography, Augmented Reality, and graphical user authentication (GUA) subjects. He also works on wireless security, mobile communication, and optical network bandwidth allocation systems as a research assistant. He is advisor of Limkokwing University of Creative technology computer and mobile research Labs.

Evercookies : Extremely persistent cookies

Mohd. Shadab Siddiqui¹
shadab-siddiqui@hotmail.com

Deepanker Verma²
deepanekr70@gmail.com

UPTU : CS&E,
SRMCEM
Lucknow, India

Abstract— Now a days, internet marketing is on peak. Most of the companies use internet for marketing and advertisements of their business. For tracking visitors, cookies are stored in their browser by these company websites. But what, if visitor deleted the cookies stored? So it is really difficult to maintain a 100% correct record of visitors. So a new type of cookie is created which is really difficult to delete. The main purpose of this is to identify a returning visitor even if he/she has tried to delete the cookies stored in the browser. Samy Kamkar created a java script API for creating persistent cookies. This API is called evercookie.

Keywords- evercookie, persistent cookie, advance visitor tracking.

I. INTRODUCTION

Evercookie is used as a advance visitor tracking mechanism. Cookies can be deleted by the user so evercookie is used in visitors web browser which difficult to delete or remove. evercookie is a javascript API that produces extremely persistent cookies. In evercookie, cookie data is stored in several type of storage location available in user's web browser. There are more than 10 locations where data is stored. So it is not possible for a user to delete data at each location because user don not know about each location. That's why it is also called extremely persistent cookie. There is a special property of evercookie. That is recreation. It means, it can recreate cookie data of other location. If evercookie found that user has removed any type of cookie, it creates them again .IPhone is also a victim of it.

1.1 Overview of EverCookie

Evercookie is a javascript API that is used for creating extremely persistent cookies. This Javascript API is very simple to use. We will also demonstrate it with the code in later part of this paper. Evercookie is persistent cookie because it can use more than 10 locations on user's browser. Most of these locations are hard to find for deleting cookie data. When creating a cookie, it uses the following storage mechanisms when available.

1. Standard HTTP cookies
2. Local shared Objects
3. Sliverlight Isolated Storage

4. Storing cookies in RGB values of auto generating, force-cached PNG's using HTML5 Canvas tag to read pixels (Cookies) back out
5. Storing cookies in Web History
6. Storing cookies in HTTP ETags
7. Storing cookies in web cache
8. window.name caching
9. Internet Explorer userData storage
10. HTML5 Session Storage
11. HTML5 Local Storage
12. HTML5 Global Storage
13. HTML5 Database Storage via SQLite

You can see that most of the locations given above are not for a common user to know about. That's why this cookie is impossible to delete for a normal internet user. Only a person having great skill in computer and programming would be able to delete the data of all these locations. But it's also not easy for him.

Evercookie API is open source so any one can download it from the website of developers.

If a user gets cookie on one browser and then switched to another web browser for surfing. It does not mean that he will not be traced. Because both web browsers still have the Local Shared Object cookie, so the cookie will reproduce the data in both browsers. This means he will still be trackable by the advertisers website.

II. USAGE

Evercookie is a marketing tool. With this tracking tool it is possible to have unique persistent identification of a specific computer and thus it can link that uniq computer to the company. Here term computer is used as a visitor of that computer. Evercookie is Javascript API. You need to know Javascript for better use of EVERcookie. It also use jquery and swfobject API. Here is a demo code to demonstrate the cookie generation. This code will show you setting cookie data on the user's browser and getting cookie data from the user's browser. This is a demo code and change it as per your use.

```
<script type="text/javascript" src="jquery-1.4.2.min.js"></script>
<script type="text/javascript" src="swfobject-2.2.min.js"></script>
<script type="text/javascript" src="evercookie.js">
</script>
<script>
var ec = new evercookie();

ec.set("id", "12345");

ec.get("id", function(value) { alert("Cookie value is " +
value) });
function getCookie(best_candidate, all_candidates)
{
    alert("The retrieved cookie is: " + best_candidate + "\n" +
        "You can also see the data what each storage
mechanism returned " +
        "by looping through the all_candidates object.");
    for (var item in all_candidates)
        document.write("Storage mechanism " +
item + " returned: " + all_candidates[item] + "<br>");
}
ec.get("id", getCookie);

</script>
```

III. METHODS USE FOR STORING

1. PNG caching

When evercookie sets a cookie, it accesses evercookie_png.php with a special HTTP cookie, different than the one used for standard session data. This special cookie is read by the PHP file, and if found, generates a PNG file where all the RGB values are set to the equivalent of the session data to be stored. Additionally, the PNG is sent back to the client browser with the request to cache the file for 20 years. When evercookie retrieves this data, it deletes the special HTTP cookie, then makes the same request to the same file without any user information. When the PHP script sees it has no information to generate a PNG with, it returns a forged HTTP response of "304 Not Modified" which forces the web browser to access its local cache. The browser then produces the cached image and then applies it to an HTML5 Canvas tag. Once applied, evercookie reads each pixel of the Canvas tag, extracting the RGB values, and thus producing the initial cookie data that was stored.

2. Web History storage work

When evercookie sets a cookie, assuming the Web History caching is enabled, it Base64 encodes the data to be stored. Let's assume this data is "bcde" in Base64.

Evercookie then accesses the following URLs in the background:

```
google.com/evercookie/cache/b
google.com/evercookie/cache/bc
google.com/evercookie/cache/bcd
google.com/evercookie/cache/bcde
google.com/evercookie/cache/bcde-
```

These URLs are now stored in history. When checking for a cookie, evercookie loops through all the possible Base64 characters on google.com/evercookie/cache/, starting with "a" and moving up, but only for a single character. Once it sees a URL that was accessed, it attempts to brute force the next letter. This is actually extremely fast because **no requests** are made to the server. The history lookups are simply locally in JavaScript using the CSS History Kicker (a tool by samy). Evercookie knows it has reached the end of the string as soon as it finds a URL that ends in "-"

IV. PROTECTION AGAINST EVERCOOKIE

It's a new concept so not so many tools and tricks available to protect Evercookie to store data on various locations of your web browser. But many there are some methods which works effectively. Basically this is a Javascript API and a javascript code is used to store the data on the user's web browser. JavaScript is the main concept behind this. Javascript code is used to identify the removed cookie and again creation of that cookie data. So prevention of this code to be execute on user's browser is the best way to prevent against Evercookie. All tools which can prevent the Javascript can be used. These are some tools and ways which can be used as effective protection

1. **Nevercookie:** This is a freely available Firefox plugin for preventing evercookie from tracking users. This plugin extends Firefox's private browsing mode by preventing Evercookies from identifying and tracking users.
2. **NoScript:** NoScript provides an extra protection against this type of scripts storing data on various locations of web browser. This script is mainly for prevention against attacks caused by scripts but it also works good against Evercookie.
3. **Safari's private browsing mode:** This is also a good way to protect against evercookie. According to the creators of evercookie, using Safari's private browsing mode defeats the evercookie system and prevents evercookies from being used between browsing sessions.
4. **BleachBit:** BleachBit is also a nice tool and use to delete evercookie tracking in Firefox, Internet Explorer, Google Chrome, Opera, and Safari. So it can be used as an effective tool against evercookie.
5. **Better Privacy:** Better Privacy is a firefox addon which protects against not deletable or persistent

cookies. It is a good protection tool for a firefox user.

V. REMOVING EVERCOOKIE

IN GOOGLE CHROME

Create evercookies

- 1) Open a new tab, then close all other windows and tabs.
- 2) Delete Silverlight Isolated Storage
Go to <http://www.silverlight.net/>
Right click the Silverlight application (any app will do)
Silverlight Preferences > Application Storage > Delete all...
Click "Yes"
* Optionally disable "Enable application storage"
- 3) Delete Flash Local Shared Objects (LSO)
Go to the Flash "[Website Storage Settings panel](#)"
Click "Delete all sites"
Click "Confirm"
- 4) Clear Browsing Data
- Wrench > Tools > Clear Browsing Data...
- Select all options
- Clear data from this period: Everything
- Click "Clear Browsing data"

IN MOZILLA FIREFOX(v3.6)

1. In the 1st tab, open three blank tabs. Go to Samy's evercookie demo page. Click on "Click to create an ever cookie". Make sure evercookie is stored in every places except 'userData' storage (it's for IE). If needed click on 'click to rediscover cookies' several times.
2. Close the first (samy's) tab.
3. In 2nd tab, open <http://www.silverlight.net/> and delete Silverlight Isolated Storage
4. In 3rd tab, open Flash "Website Storage Settings panel" page and remove Flash Local Shared Objects (LSO)
5. Press Ctrl+Shift+Del (alternatively go to Tools > Clear Recent History). Select 'Everything' from the 'Time range to clear' dropdown and check every items from the 'Details' list and finally click on 'Clear Now' button.
6. Now go to samy's page again and verify that the evercookie is removed completely.

VI. CONCLUSION

At the end of this paper we have come to **the conclusion that** Evercookie API is created for the advance tracking of visitors of a website. This new cookie generation offers unlimited user tracking to industry and market research This creates persistent cookies on user's web browser. And Normal user do not know about this and can not be able to delete this. It's almost impossible for a user to delete each data stored by the user. This is a marketing tool and is using by many companies to track their visitors on the web. But this is not

good for privacy reasons. So many researcher are trying to find the way to prevent this cookie to be stored on the user's browsers.

VII. REFERENCES:

- [1]. <http://samy.pl/evercookie/>
- [2]. <http://jeremiahgrossman.blogspot.com/2010/10/killi-ng-evercookie-google-chrome-wo.html>
- [3]. <http://singe.za.net/blog/archives/1014-Killing-the-Evercookie.html>
- [4]. http://threatpost.com/en_us/blogs/researcher-claims-evercookie-cant-be-removed-092210
- [5]. <http://arstechnica.com/web/news/2010/09/evercookie-escalates-the-zombie-cookie-war-by-raising-awareness.ars>
- [6]. <http://www.whitehatsec.com/>
- [7]. <http://www.techrepublic.com/blog/security/the-next-front-in-the-cookie-wars-fighting-the-evercookie/4672>
- [8]. <http://www.wilderssecurity.com/showthread.php?t=282702>

Authors Profile:

1. Mohd. Shadab Siddiqui :is a CEH (Certified Ethical hacker) and ECSA (Ec-Council Certified Security Analyst) and will be completing his B.Tech in july,2011 and has found vulnerability in many high profile sites and reported them.
2. Deepanker Verma: is a java developer with too much interest in security and will be completing his B.Tech in july,2011.He is also a blogger and maintains a security blog.

Potential Research into Spatial Cancer Database by Using Data Clustering Techniques

N. Naga Saranya,
Research Scholar (C.S), Karpagam University, Coimbatore-
641021,
Tamilnadu, India.
E-mail: nachisaranya01@gmail.com

Dr.M. Hemalatha,
Head, Department of Software Systems,
Karpagam University, Coimbatore-641021,
Tamilnadu,India.
E-mail: hema.bioinf@gmail.com

Abstract— Data mining, the taking out of hidden analytical information from large databases. Data mining tools forecast future trends and behaviors, allowing businesses to build practical, knowledge-driven decisions. This paper discusses the data analytical tools and data mining techniques to analyze data. It allows users to analyze data from many different dimensions or angles, sort it, and go over the relationships identified. Here we are analyzing the medical data as well as spatial data. Spatial data mining is the process of difficult to discover patterns in geographic data. Spatial data mining is measured a more difficult face than traditional mining because of the difficulties associated with analyzing objects with concrete existences in space and time. Here we applied clustering techniques to form the efficient cluster in discrete and continuous spatial medical database. The clusters of random shapes are created if the data is continuous in natural world. Furthermore, this application investigated data mining techniques (clustering techniques) such as Exclusive clustering and hierarchical clustering on the spatial data set to generate the well-organized clusters. The tentative results showed that there are certain particulars that are evolved and can not be apparently retrieved as of raw data.

Keywords- Data Mining, Spatial Data Mining, Clustering Techniques, K-means, HAC, Standard Deviation, Medical Database, Cancer Patients, Hidden Analytical.

I. INTRODUCTION

Recently many commercial data mining clustering techniques have been developed and their usage is increasing tremendously to achieve desired goal. Researchers are putting their best hard work to reach the fast and well-organized algorithm for the abstraction of spatial medical data sets. Cancer has become one of the foremost causes of deaths in India. An analysis of most recent data has shown that over 7 lakh new cases of cancer and 3 lakh deaths occur annually due to cancer in India. Cancer has striven against near insurmountable obstacles of financial difficulties and an almost indifferent ambience, to fulfill the objectives of its founder, bringing to the poorest in the land the most refined scientific technology and excellent patient care. Furthermore, cancer is a preventable disease if it is analyzed at an early stage. There are different sites of cancer such as oral, stomach,

liver, lungs, kidney, cervix, prostate testis, bladder, blood, borne, breast and many others. There has been huge development in the clinical data from past decades, so we need proper data analysis techniques for more sophisticated methods of data exploration. In this study, we are using different data mining technique for effective implementation of clinical data. The main aim of this work is to discover various data mining techniques on clinical and spatial data sets. Several data mining techniques are pattern recognition, clustering, association, and classification. Our Proposed work is on medical spatial datasets by using clustering techniques. There are fast and enormous numbers of clustering algorithms are developed for large datasets such as CURE, MAFLA, DBSCAN, CLARANS, BIRCH, and STING.

II. CLUSTERING ALGORITHMS AND TECHNIQUES IN DATA MINING

The process of organizing objects into groups whose members are similar in some way is called clustering. So, the goal of clustering is to conclude the essential grouping in a set of unlabeled data. Various kinds of Clustering algorithms are partitioning-based clustering, hierarchical algorithms, density based clustering and grid based clustering.

A. Partitioning Algorithm

K-Means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. Fig.1. shows the K-Means algorithm is composed of the following steps belongs to centroid:

1. It classifies a given dataset through certain number of clusters (assume k clusters). These points are first group centroids.
2. Grouping is done based on the Euclidean's distance.
3. And the centroids are formed on the basis of mean value of that object group.
4. The steps 2 & 3 repeats until the centroids no longer move.

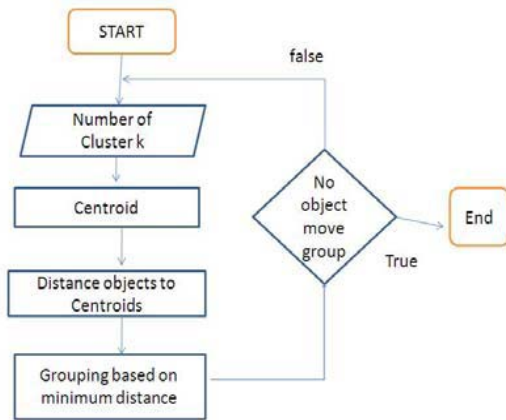


Figure 1: Work Flow of Partition based cluster algorithms

B. Hierarchical Clustering Algorithms

The hierarchical clustering functions essentially in combine closest clusters until the desired number of clusters is achieved. This sort of hierarchical clustering is named agglomerative since it joins the clusters iteratively. There is also a divisive hierarchical clustering that does a turn around process, every data item start in the same cluster and then it is divided in slighter groups (JAIN, MURTY, FLYNN, 1999). The distance capacity between clusters can be done in numerous ways, and that's how hierarchical clustering algorithms of single, common and totally differ. Many hierarchical clustering algorithms have an interesting property that the nested sequence of clusters can be graphically represented with a tree, called a 'dendrogram' (CHIPMAN, TIBSHIRANI, 2006). There are two approaches to hierarchical clustering: we can go from the bottom up, grouping small clusters into larger ones, or from the top down, splitting big clusters into small ones. These are called agglomerative and divisive clustering, respectively.

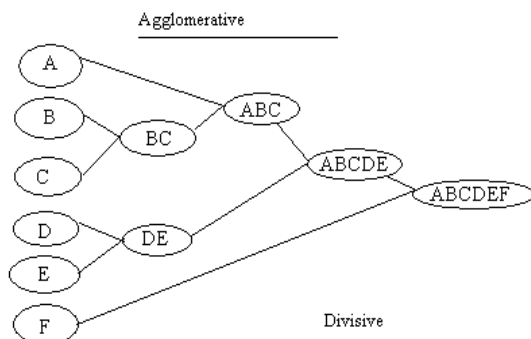


Figure 2: Hierarchical Clustering

1) *Agglomerative approach* is the clustering technique in which bottom up strategy is used to cluster the objects. It merges the atomic clusters into larger and larger until all the objects are merged into single cluster.

2) *Divisive approach* is the clustering technique in

which top down strategy is used to cluster the objects. In this method the larger clusters are divided into smaller clusters until each object forms cluster of its own. Fig.2 shows simple example of hierarchical clustering.

C. Density Based Clustering Algorithm

It is a clustering technique to develop clusters of arbitrary shapes. They are different types of density based clustering techniques such as DBSCAN, SNN, OPTICS and DENCLUE. *DBSCAN algorithm:* The DBSCAN algorithm was early introduced by Ester, et al. [Ester1996], and relies on a density-based notion of clusters. Clusters are recognized by looking at the density of points. Regions with a high density of points depict the existence of clusters whereas regions with allow density of points indicate clusters of noise or clusters of outliers. This algorithm is particularly suited to deal with large datasets, with noise, and is able to identify clusters with different sizes and shapes.

The algorithm: The key idea of the DBSCAN algorithm is that, for each point of a cluster, the neighborhood of a given radius has to contain at least a minimum number of points, that is, the density in the neighborhood has to exceed some predefined threshold. This algorithm requires three input parameters:

- k, the neighbors list size;
- Eps, the radius that delimitate the neighborhood area of a point (Eps neighborhood);
- MinPts, the minimum number of points that must exist in the Eps-neighborhoods.

The clustering process is based on the classification of the points in the dataset as core points, border points and noise points, and on the use of density relations between points (directly density-reachable, density-reachable, density-connected [Ester, 1996] [2]) to form the clusters.

D. SNN algorithm

The SNN algorithm [Ertoz, 2003] [3]) the same as DBSCAN, is a density-based clustering algorithm. The main difference between this algorithm and DBSCAN is that it defines the similarity between points by looking at the number of nearest neighbors that two points share. Using this similarity measure in the SNN algorithm, the density is defined as the sum of the similarities of the nearest neighbors of a point. Points with high density become core points, while points with low density represent noise points. All remainder points that are strongly similar to a specific core points will represent a new clusters.

The algorithm: The SNN algorithm needs three inputs parameters:

- K, the neighbors' list size;
- Eps, the threshold density;

- MinPts, the threshold that define the core points.

After defining the input parameters, the SNN algorithm first finds the K nearest neighbors of each point of the dataset. Then the similarity between pairs of points is calculated in terms of how many nearest neighbors the two points share. Using this similarity measure, the density of each point can be calculated as being the numbers of neighbors with which the number of shared neighbors is equal or greater than Eps (density threshold). Next, the points are classified as being core points, if the density of the point is equal or greater than MinPts (core point threshold). At this point, the algorithm has all the information needed to start to build the clusters.

Optics: OPTICS (Ordering Points to Identify the Clustering Structure) is the clustering technique in which the augmented order of the datasets for cluster analysis. Optics built dataset-using density based clustering structure. The advantage of using optics is it is not sensitive to parameters input values through the user it automatically generates the number of clusters.

Denclue: DENCLUE (Clustering Based on Density Distribution Function) is the clustering technique in which the clustering method is dependent on density distribution function. A cluster is defined by a local maximum of the estimated density function. Data points are assigned to clusters by hill climbing, i.e. points going to the same local maximum are put into the same cluster. The disadvantage of Denclue 1.0 is, that the used hill climbing may make unnecessary small steps in the beginning and never converges exactly to the maximum, it just comes close. The clustering technique is basically based on influence function (data point impact on its neighborhood), the overall density of data space can be calculated as the sum of influence functions applied to data points and cluster can be calculated using density attractors (local maxima of the overall density function).

E. Grid Based Clustering

The Grid Based clustering algorithm, to form a grid structure it partitions the data space into a finite number of cell. After that performs all clustering operations are obtained grid structure. It is a well-organized clustering algorithm, but its effect is greatly partial by the size of the cells. Grid-based approaches are well-liked for mining clusters in a large multidimensional space in which clusters are regarded as denser regions than their environs. The computational difficulty of most clustering algorithms is at least linearly comparative to the size of the data set. The great advantage of grid-based clustering is its important decrease of the computational complexity, especially for clustering very huge data sets [8]. In general, a distinctive grid-based clustering algorithm consists of the following five basic steps (Grabusts and Borisov, 2002) [7] :

1. Grid Structure Creation i.e., splitting the data space into a finite number of cells.

2. Cell Density Calculation for each cell.
3. Form of the cells according to their densities.
4. Identify the cluster centers according to their result.
5. Finally Traversal of neighboring cells.

III. EXPERIMENTAL RESULTS

Here we have taken several series of Datasets by using several websites and direct surveys. And we conclude applicable pattern detection for medical diagnosis.

Cancer Database (SEER Datasets): The web site called www-dep.iarc.fr/globocan/database.htm consist of datasets. It contain number of cancer patients those who registered themselves in this. The dataset consists of basic attributes such as sex, age, marital status, height and weight. The data of age group was taken from (20 - 75+) years in this group major cancers were examined. A total of male and female cases were examined for the various cancers. The data were collected and substantial distribution was found for Incidence and Mortality by Sex and Cancer site. Perhaps analysis suggests that they were more male cases those who were suffering from cancer as per opposite sex.

In this study, the data was taken from SEER datasets which has record of cancer patients from the year 1975-2008. Spatial dataset consists of location collected include remotely sensed images, geographical information with spatial attributes such as location, digital sky survey data, mobile phone usage data, and medical data. The five major cancer areas such as lung, kidney, bones, small intestine and liver were experimented. After this data mining algorithms were applied on the data sets such as K-means, SOM and Hierarchical clustering technique. The database analysis was done using XLMiner tool kit. Fig.3 represents the statistical diagram for representation between number of male and female cases for cancer.

The data consists of discrete data sets with following attribute value types of cancer, male cases, female cases, cases of death pertaining to specific cancer. They were around 21 cancers that have been used as the part of analysis. The XLMiner tool doesn't take the discrete value so it has to be transformed into continuous attribute value.

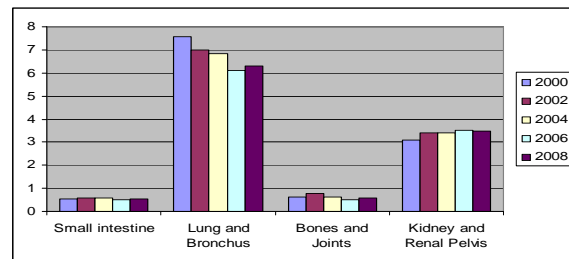


Figure 3 : Female and Male cases of Cancer

The data was subdivided into X, Y values and the result was formed using K-means and HAC clustering algorithm. In XLMiner, the low level clusters are formed using K-MEANS and SOM then HAC clustering builds the Dendrogram using the low level clusters. Fig.3 specifies the number of clusters for both sexes, in this male is more affected compare to the opposite sex.

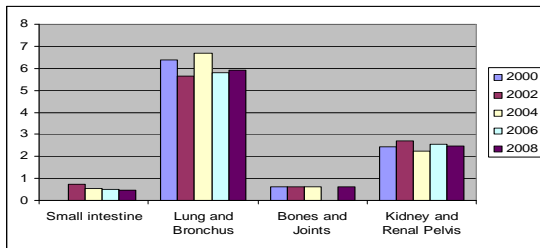


Figure 4 : Female Cases of cancer

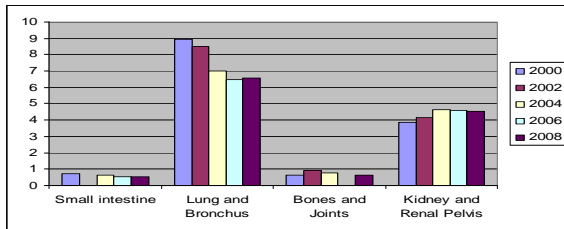


Figure 5: Male cases of Cancer

The Fig.4 and Fig.5 specifies the number of cluster for male and female suffering from different cancers. This sample is collected from the patient who couldn't stay alive with the disease. The result of the analysis shows that the male ratio was large in percentage while compared to the opposite sex. Possibly by analyzing the collected data we can enlarge certain measures for the improved procurement of this disease. . Fig.6 specifies the number of death in both males and female cases of death due to cancer using XLMiner.

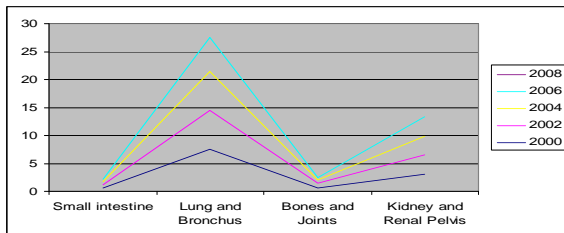


Figure 6: Female and Male death cases of Cancer

The K-means method is an efficient technique for clustering large data sets and is used to determine the size of each cluster. The input of the hierarchical clustering shown in Table1, and it contain the data, variables, parameters. These are all calculated by the distance measure which is in side the hierarchical clustering. Here X_n , ($n= 1,2,3,4,5$) are the selected variables which is placed in the datasets. After this the HAC (hierarchical agglomerative clustering), is used on our datasets in which we have used tree based partition method in which the results has shown as clustering stages and its elapsed time in Table 2.

The HAC has proved to have for better results than other clustering methods. The principal component analysis technique has been used to visualize the data. The X, Y coordinates recognize the point position of objects. The coordinates were used and the clusters were determined by appropriate attribute value. The mean and standard deviation of each cluster was determined.

TABLE1: Input of Hierarchical Clustering

Data	
Input data	['both dendrogram.xls']Sheet1!\$A\$2:\$F\$5
# Records in the input data	4
Input variables normalized	Yes
Data Type	Raw data

# Selected Variables				
Selected variables	X1	X2	X3	X4

Parameters/Options	
Draw dendrogram	Yes
Show cluster membership	Yes
# Clusters	3
Selected measure	Similarity
Selected clustering method	Euclidean distance
	Average group linkage

TABLE2: Clustering Stages

Stage	Cluster 1	Cluster 2	Distance
1	1	3	0.079582
2	1	4	1.69234
3	1	2	3.146232

Elapsed Time:

Overall (secs)	3.00
----------------	------

Table 3 presents HAC (hierarchical agglomerative clustering) in which the cluster were determined with appropriate size. Clusters are subdivided in to many sub clusters and the attributes are X_n , ($n= 1,2,3,4,5$). In this we predicted the clusters by using hierarchical clustering.

TABLE3: Hierarchical Clustering – Predicted Clusters

Row Id	Cluster Id	Sub Cluster Id	x1	x2	x3	x4	x5
1	1	1	0.5419	0.5933	0.5867	0.487	0.5273
2	2	2	7.57	6.9027	6.0376	6.1301	6.2305
3	1	3	0.6158	0.7743	0.6071	0.5157	0.5557
4	2	4	3.097	3.3904	3.3864	3.5181	3.4851

The Figure 7 represents the dendrogram in which the dataset has been partitioned into three clusters with the K-means.

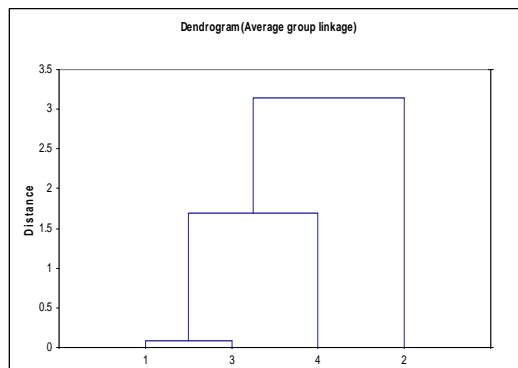


Figure 7: Dendrogram

The HAC clustering algorithm is applied on K-means to generate the dendrogram. In a dendrogram, the elements are grouped together in one cluster when they have the closest values of all elements available. The cluster 2 and cluster 3 are combined together in the diagram. Analyzes is done in the subdivisions of clusters.

TABLE 4: Representation of Cluster Mean and Standard Deviation

Cluster_HAC_1=c_hac_1				Cluster_HAC_1=c_hac_2			
Examp es		[59.3 %] 16		Example s		[7.4 %] 2	
	Test value	Grou p	Overra	Att Desc	Test valu e	Grou p	Overra
Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)			
Year ofDiagno sis	0.5	82.50 (4.76)	80.59 (24.14)	moral	2.9	64.90 (0.99)	55.78 (4.52)
Att Desc	1.7	11.79 (4.34)	13.03 (4.47)	mliver	2.7	13.60 (0.00)	6.53 (3.72)
fstomach	2.1	24.98 (5.31)	26.61 (4.84)	mstomac h	2.7	27.60 (2.97)	19.57 (4.22)
mstomac h	3.5	17.15 (1.90)	19.57 (4.22)	fliver	2.6	21.05 (1.77)	13.03 (4.47)
flungs	3.6	18.88 (1.08)	19.83 (1.64)	flungs	2.4	22.55 (0.78)	19.83 (1.64)
mliver	3.9	4.20 (1.92)	6.53 (3.72)	mkidney	2.1	70.00 (0.99)	62.13 (5.30)
mlungs	3.9	13.90 (0.68)	14.56 (1.05)	fkidney	2.1	77.75 (2.05)	66.97 (7.28)
mkidney	4	58.71 (3.83) 62.14	62.13 (5.30) 66.97	fstomach Year ofDiagno sis	1.7 0.50 4.8	32.20 (1.56) 0.50 (0.71)	26.61 (4.84) 80.59 (24.14)
fkidney	4.1	5.17 (7.28)					

Table 4 characterizes the cluster according to the mean and standard deviation of each object and cluster were determined. The primary comparison in between cluster 1 objects. The second comparison was between the objects of cluster 2 and cluster 1. The third comparison was determined in between cluster 3 and cluster 1. The results show the mean and standard deviation of each cluster and also among the objects in each cluster. The cluster 1 has the lowest number of cancer cases the cluster 2 has average number of cancer cases where as the cluster 3 has large number of cancer cases.

The cluster compactness has been determined by standard deviation where the cluster becomes compact when standard deviation value decreases and if the value of standard deviation increases the cluster becomes dispersed.

IV. Dicussion

This paper focuses on clustering algorithms such as HAC and K-Means in which, HAC is applied on K-means to determine the number of clusters. The quality of cluster is improved, if HAC is applied on K-means. The paper has referenced and discussed the issues on the specified algorithms for the data analysis. The analysis does not include missing records. The application can be used to demonstrate how data mining technique can be combined with medical data sets and can be efficiently established in modifying the various cancer related research.

V. Conclusion

This study clearly shows that data mining techniques are promising for cancer datasets. Our future work will be related to missing values and applying various algorithms for the fast implementation of records. In addition, the research would be focusing on spatial data clustering to develop a new spatial data mining algorithm. Once our tool will be implemented as a complete data analysis environment in the cancer registry of SEER datasets, we aim at transferring the tool to related domains, thus showing the flexibility and extensibility of the underlying basic concepts and system architecture.

References

1. Rao, Y.N, Sudir Gupta and S.P. Agarwal 2003. National Cancer Control Programme: Current status and strategies, 50 years of cancer control in India, NCD Section, Director General of Health.
2. Ester, M., Kriegel, H.P., Sander, J., and Xu, X. (1996) A density-based algorithm for discovering clusters in large spatial databases. In the Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD.96), Portland, Oregon, pp. 226-231.
3. [Ertöz et al. 2003] Ertöz, L., Steinbach, M., Kumar, V.: "Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data"; In Proc. of SIAM Int. Conf. on Data Mining (2003), 1-12.
4. Aberer, Karl. 2001. P-Grid: A self-organizing access structure for P2P information systems. In Proc. International Conference on Cooperative Information Systems, pp. 179-194. Springer.
5. Bar-Yossef, Ziv, and Maxim Gurevich. 2006. Random sampling from a search engine's index. In Proc. WWW, pp. 367-376. ACM Press. DOI:doi.acm. org/10.1145/135777.1135833.
6. Ng R.T., and Han J. 1994. Efficient and Effective Clustering Methods for Spatial Data Mining, Proc. 20th Int. Conf. on Very Large Data Bases, 144-155. Santiago, Chile.

7. Grabust P., Borisov A. Using grid-clustering methods in data classification// Proceedings of the International Conference on Parallel Computing & Electrical Engineering- PARELEC'2002. - Warsaw, Poland, 2002 - p. 425-426.
8. H. Pilevar, M. Sukumar, "GCHL: A grid-clustering algorithm for high- dimensional very large spatial data bases", Pattern Recognition Letters 26(2005), pp.999-1010.
9. Jones C., et al., 2002. Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT Project [C]. In proceedings: 25th ACM Conference of the Special Interest Group in Information Retrieval, pp387-388.
10. Processing of Spatial Joins, Proc. ACM SIGMOD Int. Conf. on Management of Data, Minneapolis, MN, 1994, pp. 197- 208.
11. T. Zhang, R. Ramakrishnan, and M. LInvy, B1RCH: An Efficient Data Clustering
12. Method for Very Large Databases, Proc. ACM SIGMOD Int'L Conf. On Management of Data, ACM Press, pp. 103-114 (1996).
13. M. Ester, H. Kriegel, J. Sander, and X. Xu. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", In Proc. of 2nd Int. Conf. on KDD, 1996, pp. 226-231.
14. Wang, Yang, R. Muntz, Wei Wang and Jiong Yang and Richard R. Muntz "STING: A Statistical Information Grid Approach to Spatial Data Mining", In Proc. of 23rd Int. Conf. on VLDB, 1997, pp. 186-195.
15. Ian H. Witten; Eibe Frank (2005). "Data Mining: Practical machine learning tools and techniques, 2nd Edition". Morgan Kaufmann, San Francisco.
16. M J Horner, L A G Ries, M Krapcho, N Neyman, R Aminou, N Howlader, et al. (2009) SEER Cancer Statistics Review , 1975-2007, Based on November 2008 SEER data Submission.
17. Gondek D, Hofmann T (2007) Non- redundant data clustering. Knowl Inf Syst 12(1):1-24.

Experience in teaching and published Twenty seven papers in International Journals and also presented seventy papers in various National conferences and one international conference. Area of research is Data Mining, Software Engineering, Bioinformatics, Neural Network. Also reviewer in several National and International journals.

Authors Profile

N.Naga Saranya received the first degree in Mathematics from Periyar University in 2006, Tamilnadu, India. She obtained her master degree in Computer Applications from Anna University in 2009, Tamilnadu, India. She is currently pursuing her Ph.D. degree Under the guidance of Dr. M.Hemalatha, Head, Dept of Software Systems, Karpagam University, Tamilnadu, India.



Dr.M.Hemalatha completed MCA MPhil., PhD in Computer Science and Currently working as a AsstProfessor and Head , Dept of Software systems in Karpagam University. Ten years of

AUGMENTED REALITY FOR MUSEUM ARTIFACTS VISUALIZATION

EDMUND NG GIAP WENG

Centre of Excellence for Semantic Technology and
Augmented Reality, Faculty of Cognitive Sciences and
Human Development, Universiti Malaysia Sarawak
nggiapweng@yahoo.com

LINA CHAI HSIAO PING

Centre of Excellence for Semantic Technology and
Augmented Reality, Faculty of Cognitive Sciences and
Human Development, Universiti Malaysia Sarawak

BEHRANG PARHIZKAR

Faculty of Information & Communication Technology,
LIMKOKWING University
Cyberjaya, Selangor, Malaysia
hani.pk@limkokwing.edu.my

ARASH HABIBI LASHKARI

Faculty of Information & Communication Technology,
LIMKOKWING University
Cyberjaya, Selangor, Malaysia
a.habibi.1@hotmail.com

ABSTRACT- Recently, advances in computer graphics and interactive techniques have increased the visual quality and field of Augmented Reality (AR) applications. Research into indoor exhibition systems associated with the use of AR technologies is getting general. This project describes an AR based system for overlaying computer generated information on the real world where museum artifacts are digitized in this project and superimposed in real scene. Furthermore, this project also presents the information of the artifacts in virtual form. 3D models are created and rendered in an AR environment providing an opportunity for museum visitors to visualize virtual artifacts in the context of other contextual information.

I. INTRODUCTION

An artifact is any object made or modified by a human being which contain the information about the culture of its creator and users, and later discovered by archaeologist. According to Digital Artifact (2008), the artifact may change over time in what it represents, how it appears and how and why it is used as the culture changes over time. Examples include stone tools such as projectile points, monument, pottery vessels, metal objects such as buttons or guns, and items of personal adornment such as jewellery and clothing. Other examples include bone that show signs of human modification, fire cracked rocks from a hearth or plant material used for food. Artifacts can come from any archaeological context or source such as:

- Buried along with a body (grave goods).
- From any feature such as a midden or other domestic setting
- Hoards
- Votive offerings

Besides cultural artifacts, there is also other form of artifacts. Digital artifacts is another type of artifacts which are visibly defected in a digital photo or video picture that specifically caused by the digital encoding (Digital Artifacts, 2008).

Recently virtual artifacts pop out as a new alternative for artifact. Virtual artifact is an object that not physically existing in real environment but present in digital environment, for example the Internet, virtual reality, cyberspace and augmented reality. The term "virtual artifact" has been used in a variety of ways in scientific and public discourse (Virtual Artifact, 2008). Previously it has referred to objects of different nature such as images, models, prototypes, computer animation or virtual books that exist in digital environments. Nowadays virtual artifact extends their role to different field such as museum artifacts exhibition where real world artifacts are modeled or reconstruct in a digital environment.

Museum is a place where the collections of cultural heritage is protected and exhibited. It also provides better understanding of history. According to the definition of International Council of Museums, a museum is a non-profit making and permanent institution that is open to the public. It also acquiring, protecting, researching, communicating and exhibiting cultural heritage for purposes of study, education and enjoyment (Museum, 2008).

Some recent surveys in Europe show that about 35% of museums have already started developments with some form of 3D presentation of objects (Tsapatori, 2003). Rapid development of technology results in facing the new paradigm of museum. The challenges of museum such as extended types and quantity of materials require rethinking conventional concept of museum. Even though new technologies are changing the museum environment gradually, it still keeps its role and purpose as its functions are exhibition, communication research and conservation.

One of the benefits of the advance technology for museum is it provide a tremendous way to exhibits museum artifacts. Besides conventional display in the gallery, Virtual Reality (VR) and Augmented Reality (AR) technology offer new possibilities where the artifacts are digitized to several forms so call virtual artifacts and augmented artifacts. Virtual Reality is a user interface technology that allows users to

interact with a computer simulated environment through human sensory channels in real time (Virtual Reality, 2007). The virtual world is interactive where the users interact with the system with real-time response in an effective way. Then, the users are immersed in this virtual environment.

Virtual reality technology has already reached the level of maturity which allowed it to be applied in real life application such as cultural heritage. The virtual museum and its artifacts is a generation of three-dimensional scene by computer and it requires high performance computer graphics system to provide a sufficient level of realism to the environment.

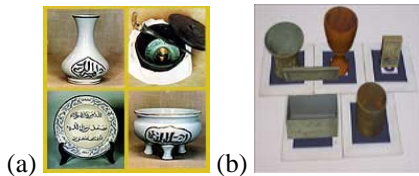


Figure 1: (a) Image of real artifacts (b) Image of virtual artifacts

Augmented reality (AR) is one of the variations of virtual reality. It is a different application where the virtual objects superimposed upon in the real world environment and the users are interacting with this virtual object in real time (Vallino, 1998). Augmented Reality on the other hand extends VR systems with the support for blending real and virtual elements into seamless composite scenes. It offers a natural view of virtual objects in real scenes. AR enhances the physical reality by integrating virtual objects into the physical world which become an equal part of the natural environment (Augmented and Mixed Reality, 2000). As stated by Vallino (1998), one of the objectives of Augmented Reality is to augment the real world while maintaining users' sense and feel of existing in a real world. In this project, real museum artifacts are converted to augmented artifacts. This project constitutes the concept of augmented museum with real museum where the exhibited artifacts are digitized to render in the augmented reality environment.

II. PROBLEM STATEMENT

Museum usually holds a lot of collection of artifacts which they cannot exhibit them publicly especially those unique artifacts, since the security and the preserving of the artifacts is one of the important factors. They need to take into account the risk on the nature and fragility of the artifacts. One of the disadvantages of the conventional exhibition is the interaction between the visitors and the artifacts are limited. The visitors cannot study the artifacts from different angles and in different context since the artifacts exhibited is static and protected in gallery or in the glass to prevent the visitors from touching and destroying it.

On the other hand, museum is a place that provides an opportunity for the visitors to learn about the history and cultural heritage. In conventional museum, the information of the exhibited artifacts is presented using panels or leaflets. But the disadvantage of this method is it only provides small

amount of information to the visitors since the information size is restricted by the physical area of the panel or paper. If the visitors want to know more about the artifacts they need to find the information by themselves through another media such as expert, books or internet. This is not a convenience method for the visitors since they cannot get the information directly from the panels or leaflets. Besides that, some museums might don't have enough space and resources for them to exhibit the whole collection of artifacts to the public. Thus an effective solution is needed so that they can exhibit the whole collection to the public without requires wide space.

III. OBJECTIVE OF STUDY

- The aim is to design and develop an augmented reality system for museum artifacts application.
- To design and develop an effective system for user and artifacts interaction using Augmented Reality technology
- To enhances the effectiveness of artifact exhibition

IV. LITERATURE REVIEW

Virtual Reality application in museum

Virtual reality is a computer generated system that allows user to immerse in interactive 3D environment (Ng, n.d., p.8). Many museum applications based on VRML has been developed for the web. This technology has been use to reconstruct the archaeological artifacts and historical sites. The examples of this technology in museum application are 3D Murale and The Rideau Street Chapel.

3D Murale

3D Murale is referring to 3D Measurement & Virtual Reconstruction of Ancient Lost Worlds of Europe. This project developed and using 3D multimedia tools to record, reconstruct, encode and visualized archeological ruins in Virtual Reality. It is aimed at developing a system capable of recording archaeology excavation phases using Virtual Reality techniques. Visualization of the reconstructed site is important both for the scientists to test and document their hypotheses in virtual reality as well as for the broad public to get an idea of how the ancient city could have looked like (3D Murale, 2003).

This project used 3D multimedia tools to measure, reconstruct and visualize archaeological ruins in virtual reality using a test case which is the ancient city of Sagalassos in Turkey. Media and textual information about archaeological content is stored in a database. This content is structured by metadata information. Metadata information will make this content available by remote Internet access through the use of search engines for archaeological researchers and members of the public. Furthermore, the project offers the reconstruction of excavated remains of pottery, sculptures and buildings as well as their visualization in a way as they possibly looked like throughout ages (3D Murale, 2003).

In this system, the users are able to freely navigate the ancient city. Collision detection is added in this system to

avoid users passing through objects. To add the educational value, various kind of interactive multimedia such as textual information, images, movie, sound and recorded objects is integrated into the virtual environment. Guided tours is also provided to guide users navigate through the system. User can select to use signpost or an avatar virtual guide instead.



Figure 2.1 Guide tours through signpost (left) and virtual guide (right).

The Rideau Street Chapel

The Rideau Street Chapel or chapel of the *Convent of Our Lady of the Sacred Heart* in Ottawa was demolished by a developer in 1972. Luckily, its architecturally unique interior was taken apart and later reconstruct inside the National Gallery of Canada where it is currently preserved. Reconstructing a historical site as it once was or as it evolved over time is one of the most important goals of virtual heritage (El-Hakim, MacDonald, Lapointe, Gonzo & Jemtrud, 2006).

This project objective is to digitize and model the existing interior and reconstruct the destroyed exterior from old images and drawings to create a complete virtual reconstruction of the chapel as it once was. The steps applied for modeling and visualization of a heritage site through time is CAD modeling from existing engineering drawings, laser scanning with two different scanners, Photogrammetry, and modeling from old photos (Figure 2.3 and Figure 2.4). The existing engineering drawings, which were based on surveying and Photogrammetry, created the overall model of the interior of the chapel. The outside of the chapel was modeled from photographs taken before 1972. All models were integrated together and presented with the tools.



Figure 2.2: Rideau Chapel: old images (up), virtual model (down)

After that, all 3D models and other data are assembled by linking components to each other, correcting scale, filling gaps, and creating smooth transitions. An interactive presentation and high quality pre-rendered animations is created with all models and data and light modeling is done with different light types at various daytimes and seasons to increase the realism of the model.

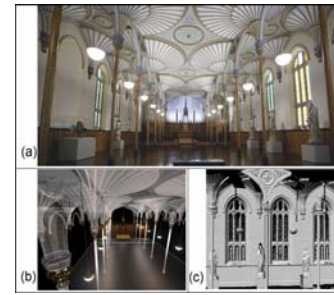


Figure 2.3 Current chapel interior: (a) overall view, (b) part of wire-frame model, (c) scanned data

Augmented Reality

Augmented reality (AR) can be defined as referring to cases in which a real environment is “augmented” by means of virtual objects (Milgram & Kishino, 1994). AR has a wide variety of uses, as it can clearly demonstrate spatial concepts, temporal concepts and contextual relationships between both real and virtual objects (Aldridge, Bilinghurst, Garrie & Woods, 2004).

Figure 2.4 shows Milgram’s Reality Virtuality Continuum which describes the relationship between augmented reality and virtual reality (Milgram & Kishino, 1994). Augmented reality lies near the real world end spectrum with the perception in the real world augmented by computer-generated data. Augmented reality is a system which most objects in the environment are synthetic with some real world images mapped on it.

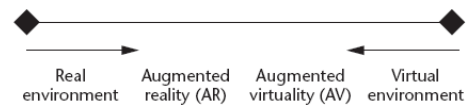


Figure 2.4 Milgram Reality Virtuality Continuum

Thus augmented reality technology creates partially virtual and real environments which enable it to be applied to different application such as medical, entertainment, education, robotics or architectures. The goal of augmented reality systems is to combine the interactive real world with an interactive computer-generated world in such a way that they appear as one environment (Vallino, 1998).

Augmented Reality Application in Museum

Augmented Reality is a promising technology that can have wide impact on cultural heritage. A museum for example is one of the best places for AR applications. This is because they offer many challenges to AR researchers such as finding novel ways of providing information and offering new consultation methods for archaeological or cultural sites (Liarokapis & White, 2005). Using this technology, the artifacts are digitized to replace the original artifacts in the exhibition and offer more interactive method.

Since past few years, Augmented Reality is beginning to be used in museum application. AR offers an interface technology that aims to exploit ways of combining computer-generated information with the real world. User in AR environments can interact in a completely natural way

unlike with virtual reality technology, which immerses participants in a completely synthetic environment. The use of AR technology in museums promises great advances in natural interaction with museum artifacts and their data. Usually the visitors expect the visualized information to be naturally presented in an instinctive and entertaining manner. Augmented reality technology has been applied in museum project like ARCO and Tokyo Digital Museum.

ARCHEOGUIDE

ARCHEOGUIDE is a project aim to design and develop a system that will fundamentally change the way visitors view and learn about a cultural heritage site. This project provide access to information in cultural heritage sites through the development of a system based on advance techniques including Augmented Reality, 3D visualization, mobile computing and multi-modal interaction (ARCHEOGUIDE, 2001). ARCHEOGUIDE provides an interactive AR guide for the visualization of outdoor archaeological sites.

In this project, visitors are provided with a wearable computer equipped with a Head Mounted Display (HMD), camera and speaker and a lightweight portable computer. According to Hildebrand et al. (2001), the system guides the visitors through the site, acting as a personal intelligent aid giving them audio and visual information. For example, real views are enhanced with 3-D reconstructions of ruins and artifacts, avatar videos, and audio commentary to help them gain more insight into relevant aspect of the site.



Figure 2.9 Natural view (left) and Augmented view (right).

ARCHEOGUIDE is built on server-client architecture. It consists of 3 main functional blocks: Site Information Server, Mobile Units and Wireless Network. The site information server is a powerful multi-processor system that used multithreading to support many concurrent user requests for I/O or processing (Hildebrand et al., 2001). The Mobile Unit (MU) is essentially a wearable computer consisting of a laptop computer and a see-through Head Mounted Display (HMD) equipped with earphones, speaker and a camera that sends its input, almost the same as the user's optical field to the frame grabber of the wearable computer.

A wireless network with a sufficient number of Access Points (AP) provides connectivity to the server who is responsible for updating the contents of the MU's database whenever the user is moving to an area about which there is no content available (Gleue & Dähne, 2001). The MU also contains a Global Positioning System (GPS) that will be used as a first order position tracker. GPS serves two

purposes which are to restrict the number of recognizable landmarks the system searches for in the camera image, and to maintain at least an approximate position tracking information at all times (Hildebrand et al., 2001).

In this project, 3D Studio and Alias Wavefront is used to model the architects for the 3D models as shown in Figure 2.9. For image rendering, VR toolkit which is Avalon is used because it support 3D output devices like HMD and input devices like 6D tracking system and also media such as live video and audio. Furthermore it allows putting stereo textures onto objects (Hildebrand et al., 2001). These features make it suitable to use for AR application.

ARCO

The Augmented Representation of Cultural Objects Project (ARCO) aimed at develops technology, system and expertise museums need for digitization and visualization of cultural heritage artifacts. It develops the whole chain of technologies to help museums to create, manipulate, manage and present digitized cultural objects in virtual exhibitions accessible on the web (ARCO, 2003).

In this project, 3D models generated by the object modeler are refined in the interactive model refinement and rendering tool. As stated in ARCO (2003), virtual representations and associated metadata are then exported to the XML driven along with associated digital photographs, 3D models, and descriptive metadata. Virtual exhibitions are then dynamically visualized using X-VRML templates in several ways through a VRML or X3D based web browser which enable the system works over the Internet, or on a touch screen display in the museum or through a Table-top Augmented Reality Environment.

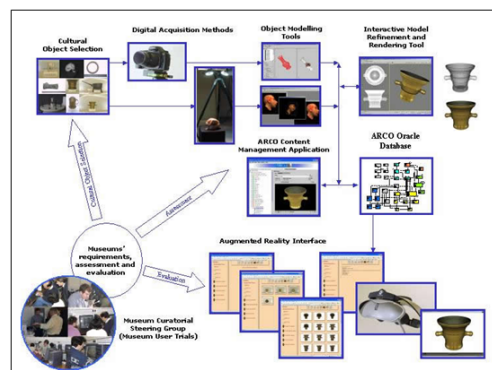


Figure 2.10 ARCO project data flow (ARCO, 2003).

In the Web-based interface a user can browse information presented in a form of 3D VRML virtual galleries or 2D Web pages with embedded multimedia objects. The Web-based interface requires a standard web browser such as Internet Explorer with a VRML plug-in. This kind of user interface can be used both on local displays inside a museum and remotely on the Internet. Augmented Reality interface refers to the interfaces dynamically generated with X-VRML that allow visualization of the database content both within museums and on the web. Two main variants of

the augmented reality interface are planned, a web browser which is either locally or remotely and an AR browser. Both variants allow the user to visualize an artifact through a set of media such as VRML, QuickTime object movies and so on (White et al., 2004).

For the AR visualization, a camera and a set of physical markers placed in a real environment is used. Video captured by the camera is passed on to the AR browser that overlays virtual representation of virtual museum artifacts using the markers for object positioning. The content and layout of the visualized scenes are determined by visualization templates that define which components of a virtual museum artifacts are composed into one VRML scene (White et al., 2004). Users can interact with the displayed objects using both the markers and standard input devices, such as the Space Mouse. A user can manipulate a marker in front of a camera and look at overlaid objects from different angles and distances. This is a natural and intuitive method of interaction with virtual objects (ARCO, 2003).

Tokyo Digital Museum

As stated by Koshizuka (n.d.), one of the main objectives of "Museum" is to provide visitors a chance to learn about the cultural heritage. For this objective, information-providing techniques for visitors are very important. Thus, Tokyo University Digital Museum, have started developing a new information-providing system for exhibited materials by using AR technologies in the computer science field. Differ with Virtual Reality technology, Augmented Reality technology realize some kinds of un-natural and artificial magic in the real world (Koshizuka, n.d.).

In this system, a user can see real images through head mounted display (HMD). HMD contains an LCD screen in front of the user's eyes, which can display information from a computer over the real images. This device includes a small camera, which measures distance between the visitor and materials, and recognizes what the user is looking at (Gleue & Dähne, 2001). In addition, the user needs to bring an electronic tag that embedded in many places for more flexible information providing, this help to keeping track of the user's position in the exhibition room (Koshizuka, n.d.). This system can provide unrestricted and large amount of information to user.

As stated by Koshizuka (n.d.), user can see information of the same materials, and large panels or large devices are not required. Thus the advantages shown by this system is the user able to see the explanation or information of exhibition in front of their eyes and listen to the aural information at the same time while traveling around the museum.

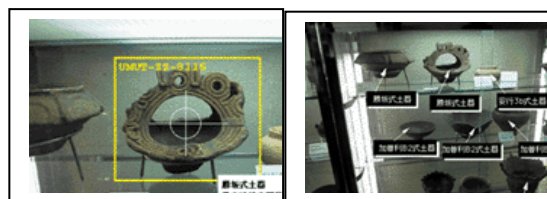


Figure 2.11 Virtual texts in real scene (Koshizuka, n.d.).

3D Modeling software

A 3D modeler is a 3D computer graphics software application used to visually produce polygonal 3D models, or the title of a professional who uses the software to produce 3D models. The 3D model is most often described by a geometric mesh of triangles or other polygons located in 3D space, with other popular descriptive models such as subdivision surface models (3D Modeling Software, 2008). There are many modeling software in market with its own specialty. Some are specially designed to model certain objects, such as chemical compounds or internal organs. 3D modelers contain a number of related features, such as rendering alternatives and texture mapping facilities. Some also contain features that support or allow animation of models. Some may be able to generate full-motion video of a series of animation.

3D modelers allow users to create and alter models via their 3D mesh. Users can add, subtract, stretch and otherwise change the mesh to their desire. Models can be viewed simultaneously. Models can be rotated and the view can be zoomed in and out. Many 3D modelers are general-purpose and can be used to produce models of various real-world entities, from a single stone to machine to human. Furthermore, 3D modelers can export their models to other applications. This is because many modelers allow importers and exporters to be plugged-in, so they can read and write data in the native formats of other applications (3D Modeling Software, 2008).

SolidWorks

Solidworks is used primarily by mechanical engineers and designers. SolidWorks is a Parasolid-based solid modeler, and utilizes a parametric feature-based approach to creating models and assemblies (SolidWorks, 2008). Parameter is a limit or boundary which defines the scope of a particular process or activity which determine the shape or geometry of the model or assembly. Parameters can be either numeric parameters, such as line lengths or circle diameters, or geometric parameters, such as tangent, parallel, concentric, horizontal or vertical.

As stated in SolidWorks (2008), building a model in this software usually starts with either a 2D or 3D sketch. The sketch consists of geometry such as lines, arcs, conics, and spine. Relations are used to define attributes such as tangency, parallelism, perpendicularity, concentricity, and such. The parametric nature of SolidWorks means the dimensions and relations drive the geometry. User can control the dimensions in the sketch and related it by relationships to other parameters outside the sketch.

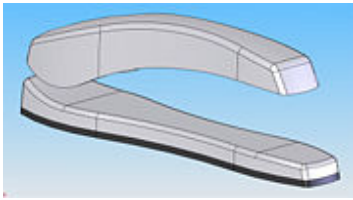


Figure 2.12 Model built using SolidWorks.

3dsMax

3ds Max is a full-featured 3D graphics application developed by Autodesk Media and Entertainment. It is the most widely-used off the shelf 3D animation program by content creation professionals according to the Roncarelli report. It has strong modeling capabilities, a flexible plugin architecture and a long heritage on the Microsoft Windows platform. This software is mostly used by video game developers, TV commercial studios and architectural visualization studios. It is also used for movie effects and movie pre-visualization (3ds Max, 2008).

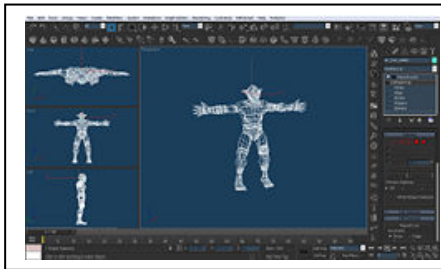


Figure 2.13 3ds Max Version 2008 interface with a dark UI.

MilkShape 3D software

There are plenty of great 3D packages on the web with proprietary and open source 3D file formats. MilkShape 3D is one of the easier to learn or more affordable modeling software. Therefore, artifacts of this project are modeled using MilkShape 3D software. MilkShape 3D (MS3D) is a low polygon 3D modeling program created by Mete Ciragan. It is used mainly by people compiling models for Half-Life and other games. MilkShape 3D allows the easy creation and manipulation of polygonal characters and objects through the use of tools which are both intuitive and easy to access. It is also a skeletal animator. This allows exporting to morph target animation like the ones in the Quake model formats or to export to skeletal animations like Half-Life, Genesis3d or Unreal.

Besides that, this software has all basic operations like select, move, rotate, scale, extrude, turn edge, subdivide and it also allows low-level editing with the vertex and face tool, standard and extended primitives like spheres, boxes, cylinders, and so on (Ciragan, 2006).

V. SYSTEM ARCHITECTURE

This project has three main components involve in the architecture of the system which are content production stage, database stage and content visualization stage.

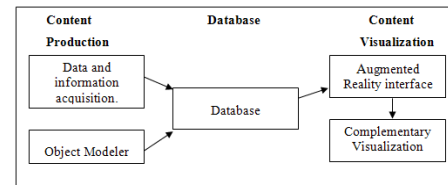


Figure 3.1 Systems architecture.

Content production stage will responsible to collect all necessary data required for digitization such as movies or images of artifacts. In this project, the image and information of the artifacts are acquired from Museum Negeri Sarawak and internet sources. Once the information is gathered, the data will be stored in filing system for further processing. Besides that, this stage also includes all tools and techniques used to create digital model of museum artifacts. To accurately model the shape, a modeling tool is needed in this project. Thus, MilkShape 3D software has been chosen as a tool to digitize the artifacts. Digitized models are then require further content refinement such as reconstruction of missing parts or repair of polygon meshes. Repair operations such as merging meshes and eliminating overlapping polygons are carried out.

In database management process, the information and digital artifacts model created is stored in the database. Each of the digitized artifact models is represented as a set of associated metadata and media object for this project. Example of media objects are images, 3D models and detail of the artifacts. The database is the central component in this system because it stores, organizes and manage digitized artifacts model into collections for the exhibition.

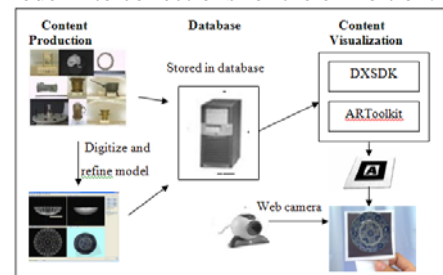


Figure 3.2 System architecture diagram

Finally in last stage, the software and toolkit such as Microsoft Direct X and ARToolkit are installed to generate the system. The digitized artifacts are input to the ARToolkit which fully control the visualization development. ARToolkit overlays digitized artifacts upon video frame captured by a camera which will gives user an impression that the virtual artifacts actually exist in the real environment. The digitized information is seamlessly superimposed into the real world in a complementary way. The implementation of the AR visualization scenarios are based on the software and hardware components. For instance the web camera is used to capture the image of the tabletop environment and a workstation with the display monitor to visualize the superimposed information. In this

stage, the digitized artifacts model and its information will pop out once the marker is detected by the system.

VI. MODEL SELECTION

In this phase, artifacts model are selected from the two resources. The artifacts are obtained from Museum Negeri Sarawak and internet resources. Through the analysis, suitable tool and model is selected to convert to digitized model.

Artifacts modeling

As mentioned before, the artifact models are digitized to 3D model using modeling program which is MilkShape 3D software. In order to build 3D artifact model that accurately representing the original model, there are few steps need to be take into caution. For example, correct mesh and accurate angle.

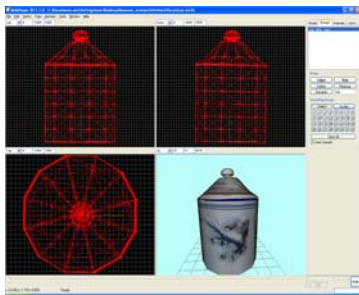


Figure 4.1 Final model of China Cup.

Figure 4.3 show an example of the model. First, a suitable shape is selected from Model bar. For example, cylinder shape and scale tool is chosen to build the body of the China Cup. This process continues until a complete model of *China Cup* is created. Then, the model is check with original model to refine to repair the polygon meshes. Once the model is done, the modeling process is continuing with texturing the model. Using the texture created, the texture applied to the model using the functions provides in Material bar. Finally the model is saved as 'ms3d' format.



Figure 4.2 Material bar of MilkShape 3D software

Texture capturing

After selecting the suitable model, its texture is captured using image editing program. Before the texture is created, the wireframe of the model is obtained using LithUnwrap software. This will make the texturing work easier since we will be able to know the shape and size of the model.

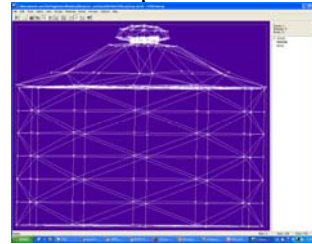


Figure 4.3 Wireframe of the model from LithUnwrap.

In this project, Adobe Photoshop 6.0 is chosen to create the texture of the artifacts from the artifact's image selected.

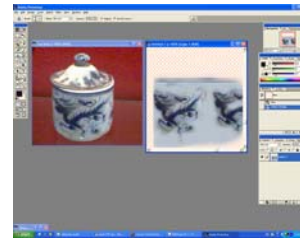


Figure 4.4 Using Photoshop tool to build the texture.

Image size of 256 * 256 pixels and transparent background is chosen for the texture. Then apply the wireframe obtained from LithUnwrap software as second layer. After that Photoshop tools such as clone tool and paintbrush tool is used to create the texture from the original image. Figure 4.5 shows the original image and its texture creates using Adobe Photoshop 6.0.

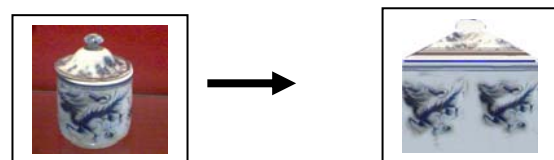


Figure 4.5 Image of *China Cup* and its texture.

VII. CODING PHASE

In this section, the coding used to implement the system will be deeply discussed. Since this system is developed from Ng's ARToolkit prototype, most of the coding are not been modified.

Load Image

The image loader for this system is TextureBMP.cpp and MyTexture.cpp. These files load bitmap image with 256 * 256 pixels size and then convert it to texture. In this algorithm, the bitmap image is loaded according to its filename.

```
AUX_RGBImageRec *LoadBMP(const char *Filename)
{
    FILE *File=NULL;
    if (!File)
    {
        return NULL;
    }
    File=fopen(Filename,"r");
    if (!File)
    {
        fclose(File);
        return auxDIBImageLoad(Filename);
    }
}
```

Figure 4.6 Coding for Bitmap image loader.

Mapped texture

The surface color or texture of the model must be recreated separately in a very labor-intensive process. This is currently achieved by taking the image of the object and applying them to the surface of the virtual object as texture maps. To map the texture, the artifacts model is called and load with their specific texture. From the database, the data of the texture is retrieved and assigned to the model through gluBuild2DMipmaps function and Draw function. The sample coding used to associate the texture is as shown in Figure 4.7. Digital image file format for the texture is 'bmp' format and the filename of the texture is 'plate1' where it is located in folder 'plate'. The texture is then map to the model named 'FlowerDish' using Draw function.

```
TextureImage[0]=LoadBMP("ms3d\\plate\\plate1.bmp");
glEnable(GL_TEXTURE_2D);
gluBuild2DMipmaps(GL_TEXTURE_2D, GL_RGB, 256, 256, GL_RGB,
GL_UNSIGNED_BYTE, TextureImage[0]->data);
g_ms3d[0].Draw("ms3d\\plate\\FlowerDish.ms3d");
```

Figure 4.7

Coding of texture mapping.

Information display

Since AR can merge the virtual model with real scene, virtual text should also have the same utility. The virtual text is the computer generated text that augments to the system. The artifact's information will display in real scene when the marker is detected. The coding used is as follow:

```
glRasterPos2f(x, y);
For (i = 0; i < len; i++)
{
    glutBitmapCharacter(font, string[i]);
}
```

Figure 4.8 Coding of Virtual texts.

The information was displayed in Times New Roman with 24 points size. The function used is bitmap_output which contains four parameters (x coordinate, y coordinate, string and font). The text displayed in x and y coordinated using glRasterPos2f function. As artifact information display is kept separate from the artifact models, it is easier to author and maintain; descriptions can be changed without loading a 3D model.

```
bitmap_output(x coordinate, y coordinate,
string, font);
```

Figure 4.9 Coding for bitmap_output.

Keyboard input

In order to rotate the 3D artifacts, user needs to press some key on the keyboard. The coding used to associate with the keyboard input is shown below. So when key '1' is pressed, the 3D artifacts will rotate in x direction and when key '2' is pressed the 3D artifacts will rotate in y direction and in z direction if key '3' is pressed. This function also enable user to see the 3D artifact to rotate in two directions by pressing two assigned key. As example the 3D artifact will rotate in xy direction when key '1' and '2' is pressed.

```
If (key '1' is pressed)
{
    Rotate is true;
    Rotate at x direction;
    Increase speed;
}
If (key '2' is pressed)
{
    Rotate is true;
    Rotate at x direction;
    Increase speed;
}
If (key '3' is pressed)
{
    Rotate is true;
    Rotate at x direction;
    Increase speed;
}
```

Figure 4.10 Algorithm for keyboard input.

VIII. DISCUSSION

In this project, a system is developed to digitize museum artifact to virtual artifact using ARToolkit, OpenGL and other soft and hard components. As a result, the system is successfully developed based on Ng's ARToolkit prototype. In this system, virtual text is displayed together with the virtual artifact so that user can get the information of the artifact such as name and where it is from directly from the model. As shown in Figure 5.1, the information of the artifact is displayed on top of the artifact whereas virtual text at the bottom of the artifact shows the instruction to rotate the artifact. Using the above functions, user can interact with the content in a more intuitive and exciting manner. They can interact with the artifacts in real time such as study them in different contexts without the existing of interference as in conventional exhibition method. User can choose the artifact they prefer using the magic book presented. Inside the magic book, different marker that represents different artifacts is provided. User can experience the augmented exhibition of their preferred artifact by open the page that contains their preferred artifact.



Figure 5.1 Screen shot of virtual artifact.

Table below show the outlined of the objective achieved.

Table 2: Achievement of objectives

Objectives	Achievement
Design and develop an effective system for user and artifacts interaction using Augmented Reality technology.	System is developed successfully using Ng's ARToolkit prototype.
To enhances the effectiveness of artifact exhibition	Few feature such as virtual text and keyboard interactions are added to increase the interaction performance. This solution has solved the limitations of conventional museum exhibition.

A technical evaluation is conducted to detect the usability of the system. Few criteria of this system are tested such as rendering speed, detection speed, and sharpness of the model and so on. There are four tasks have been conducted to test the system. These tests are done to see the effects on the virtual model and text under different circumstances. Those tests are:

1. Turn the marker and web camera.
2. Place the web camera on different heights.
3. View the model from different position. (up, down, right, left)
4. Use different camera resolution. (160*120, 320*240, 640*480)

For test 1, when the marker is turn in different orientation such as x direction, there was no effect on the virtual artifact. The 3D model is still superimposed on the marker and appears as long as the web camera detected it. The same situation also happens when web camera is turned in different orientation. The only one affected was the virtual text. When web camera turn in different orientation, the virtual text will follow the movement of the camera and turn in same orientation. This situation might caused by the non-static characteristic of virtual text.

From the result, it shows that there is an effect on the distance of camera location on the virtual text but no effect for the virtual artifact. Figure 5.2 below shows the result of this test. This test is done using camera resolution of 320*240 pixels size. From the image, we can see that virtual text's spacing is getting narrower as web camera's height increase.



Figure 5.2 Web Camera placements on different heights (highest to lowest).

In test 3, web camera detects the marker in four directions which are up, down, right and left. Same as above tests, virtual artifacts shows that there was no effect on it whereas the virtual text shows some advantages of its non-static feature since the text still can still be clearly viewed by the user. The rendering speed for both virtual models is same and they appear when they have been detected.

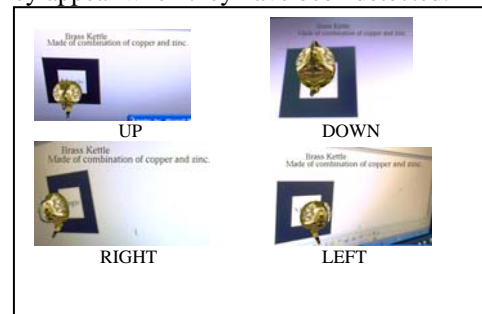


Figure 5.3 Screen shot of virtual models on different directions (Camera resolution: 320*240 pixels)

For last test, the web camera with model Logitech QuickCam Pro 5000 is used to test its resolution on the system. The resolutions tested are 160*120 pixels, 320*240 pixels and 640*480 pixels. Same as before, the effects appear on virtual text where it's spacing on screen was changing according to the resolution of the web camera. The projection of virtual text need to be improved to avoid same case happens again.



Figure 5.4 Screen shot of different camera resolution.

As a result, the evaluation done shows that there are still some limitation on this system. Virtual text in this system is non-static where it has cause some problem when the web camera is placed on different heights, using different camera resolutions or when the camera is turned in different orientation. The problem in test 1 does not affect the performance of the system since keyboard interaction has been applied in this system. Rendering speed and detection speed of this system is very satisfying since the object will pop out when the camera is detected. From the result, we can determine that web camera with resolution 640*480 pixels perform best result where the virtual text can be view clearly and sharpness of the model is better than other resolution tested.

IX. STRENGTH AND WEAKNESSES

Although there are some limitations in this system, it still shows its strength in some conditions:

- This project does not require expensive equipment such as HMD used in ARCHEOGUIDE project. It can increase the interaction between visitors and museum artifacts and also improve the exhibition performance by using standard display hardware which is a web camera and a PC set. This is more cost effective method and suitable for indoor exhibition.
- Virtual text and virtual artifact will appear as soon as the marker is detected. This increased the exhibition performance.
- Virtual text can be displayed overlay with the user's scene of view. This is corresponding to the objective stated by Koshizuka (n.d.) where information-providing techniques are very important for visitors to learn about the cultural heritage. Effective techniques in presenting the information will enhance visitor's enthusiasm to learn.
- Virtual model appear on the position set by developer and follow accurately user's instruction. They will rotate in the direction as ordered by user.
- Museum artifacts were successfully converted to 3D model where the virtual model provides a sufficient level of realism to the original model using MilkShape 3D model.

As discussed on previous section, there are some limitations in this system. From the result of the evaluation, it seems that virtual text caused the main problem. The projection of the virtual text on screen caused it to move according to web camera movement. It will turn as web camera is turned or caused difficulties in reading the text when different camera resolution is used and when the web camera is placed too far from the marker. However there is an advantage on this limitation; visitors can always see the virtual text facing them when they turn the marker or web camera in any direction as shown in Figure 5.3.

Another flaw occurs when the user manipulates the marker cards quickly, causing the computer-generated information to disappear from their view. This is due to the dynamic

registration problem. Another limitation of this AR system is the poor effectiveness of the tracking algorithms in bright lighting conditions. When this situation happen, it is difficult for the web camera to detect the shape or icon on the marker. This is because the lights might create glare spots on the marker that cause the system difficult to find the marker square.

X. CONCLUSION

Generally this system has successfully achieved its objective as states in chapter one. This system proposed a new alternative to present museum artifact. The presented solution enables museum to exhibits their artifact collections using different form of data representation which through Augmented Reality technology.

The limitation of conventional artifact exhibition is solved. Using this system, the interaction between visitor and the artifact is increase and indirectly their enthusiasm on learning about the cultural heritage could be evoked and increased. Problem such as limited space for exhibition or preserving issue in exhibiting unique artifacts is no longer a problem for museum. Furthermore, this presenting method is more interesting if compare to conventional method.

Apart from that, this system also works according to the system design. However, there are still insufficient in this system. From the technical evaluation, some improvement can be made on the algorithm to solve these limitations.

Future work

Future works for this system should include the design of advanced visualization templates for presenting cultural objects in a more attractive and appealing way. Interface will attract more users to test and used the system. Therefore, interesting interface should be design on this system. Furthermore, some other features of this system can be improve by adding in more functions such as using finger tracking method to replace keyboard. This is because finger tracking is a more natural way compared to device apparatus. Apart from that, the possibility of using haptic device in respect to the usability could also be investigated. Using this device, user can feel the artifact when they 'touch' the artifact. They can sense the interface of the artifacts or its weight. This will increase the realism of the artifacts.

REFERENCES

- [1] 3D Modeling Software (2008). *Wikipedia*. Retrieved April 2, 2008, from http://en.wikipedia.org/wiki/3D_modeler
- [2] 3D Murale (n.d.). Retrieved July 25, 2007, from <http://dea.brunel.ac.uk/project/murale.html>
- [3] 3ds Max (2008). *Wikipedia*. Retrieved April 2, 2008, from http://en.wikipedia.org/wiki/3D_Studio_Max

- [4] Aldridge, G., Bilinghurst, M., Garrie, B., & Woods, E. (2004). Augmenting the Science Centre and Museum Experience. Retrieved October 24, 2007, from <http://delivery.acm.org/10.1145/990000/988873/p230-woods.pdf>
- [5] ARCHEOGUIDE (2001). Augmented Reality-based Cultural Heritage On-site Guide. Retrieved October 21, 2007, from <http://archeoguide.intranet.gr>
- [6] ARCO (2003). Augmented Representation of Cultural Objects. Retrieved August 8, 2007, from <http://www.arco-web.org>
- [7] Artifact (2008). *Wikipedia*. Retrieved April 2, 2008, from <http://en.wikipedia.org/wiki/Artifact>
- [8] Augmented and Mixed Reality (2000). *Shared Space: an overview*. Retrieved August 20, 2007, from <http://www.mic.atr.co.jp/~poup/research/ar/index.html>
- [9] Augmented Reality (n.d.). *Virtual Advertising*. Retrieved August 20, 2007, from http://www.vtt.fi/multimedia/index_iv.html
- [10] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & Macintyre, B. (2001). Recent Advances in Augmented Reality. *Computer Graphics and Application*, 01, 0272-1716. Retrieved August 8, 2007, from IEEE database.
- [11] Ciragan, M. (2006). MilkShape 3D. Retrieved September 22, 2007, from <http://www.milkshape3d.com/>
- [12] Digital Artifacts (2008). *Wikipedia*. Retrieved April 8, 2008, from http://en.wikipedia.org/wiki/Digital_artifact
- [13] DirectX (2007). *Wikipedia*. Retrieved September 22, 2007, from <http://en.wikipedia.org/wiki/DirectX>
- [14] El-Hakim, S., MacDonald, G., Lapointe, J.-F., Gonzo, L., & Jemtrud, M. (2006). On the Digital Reconstruction and Interactive Presentation of Heritage Sites through Time. Proc. of the 7th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST 2006), NRC 48782.
- [15] Gleue, T., Dähne, P. (2001). Design and Implementation of a Mobile Device for Outdoor Augmented Reality in the ARCHEOGUIDE Project, Proc. of the Virtual Reality, Archaeology, and Cultural Heritage International Symposium (VAST01), Glyfada, Nr Athens, Greece.
- [16] Hildebrand, A., Daehne, P., Seibert, F., Christou, I. T., Almeida, L., Weidenhausen, J., et al. (2001). Design and Development issues for ARCHEOGUIDE: An Augmented Reality based Cultural Heritage On-Site Guide, Proc. of the International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging, Mykonos, Greece.
- [17] Koshizuka, N. (n.d.). Museum Navigation System using Augmented Reality Technologies. Retrieved July 22, 2007, from http://www.um.u-tokyo.ac.jp/publish_db/2000dm2k/english/01/01-16.html
- [18] Lamb, P. (n.d.). ARToolkit. Retrieved July 28, 2007, from <http://www.hitl.washington.edu/artoolkit/>
- [19] Liarokapis, F., & White, M. (2005). Augmented Reality Techniques for Museum Environments. Retrieved July 30, 2007, from <http://www.soi.city.ac.uk/~fotis/publications/MEDJCN2005.pdf>.
- [20] Microsoft Visual C++ (2007). *Wikipedia*. Retrieved September 22, 2007, from <http://en.wikipedia.org/wiki/C++>
- [21] Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Display. *IEICE Transactions on Information Systems*, Vol E77-D, No. 12. Retrieved August 3, 2007, from IEEE database.
- [22] Milgram, P., Zhai, S., Drascic, D., & Grodski, J. J. (1993). Applications of Augmented Reality for Human-Robot Communications. *RSJ International Conference on Intelligent Robots and Systems*. Retrieved August 3, 2007, from IEEE database.
- [23] Museum (2008). *Wikipedia*. Retrieved April 2, 2008, from <http://en.wikipedia.org/wiki/Museum>
- [24] Ng, G. W. (n.d.). Virtual Reality. *What is VR?*
- [25] OpenGL (2007). *Wikipedia*. Retrieved September 22, 2007, from <http://en.wikipedia.org/wiki/openGL>
- [26] SolidWorks (2008). *Wikipedia*. Retrieved April 2, 2008, from <http://en.wikipedia.org/wiki/SolidWorks>
- [27] Tsapatori, M. E. (2003). ORION Research Roadmap for the European archaeological museums' sector. Retrieved July 30, 2007, from <http://www.orion-net.org/>
- [28] Vallino, J.R. (1998). Interactive Augmented Reality. Retrieved 28 February, 2007, from <http://www.se.rit.edu/~jrv/publications/VallinoThesis.pdf>

- [30] Virtual Artifact (2008). *Wikipedia*. Retrieved April 2, 2008, from http://en.wikipedia.org/wiki/Virtual_artifact
- [31] Virtual Reality (2007). *Wikipedia*. Retrieved August 18, 2007, from http://en.wikipedia.org/wiki/Virtual_reality
- [32] Wagner, D., & Schmalstieg, D. (2007). ARToolkitPlus for Pose Tracking on Mobile
- [33] Devices. Retrieved August 23, 2007, Proc. of the Computer Vision Winter Workshop 2007, ST. Lambrecht, Australia.
- [34] White, M., Mourkoussis, N., Darcy, J., Petridis, P., Liarokapis, F., Lister, P., et al. (2004). ARCO-Architecture for Digitization, Management and Presentation of Virtual Exhibition. *Computer Society*, 04, 1530-1052. Retrieved October 31, 2007, from IEEE database.

BIOGRAPHY

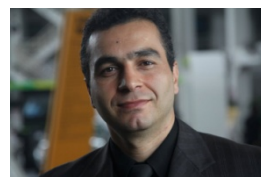


Dr. Edmund Ng Giap Weng is a lecturer at the Faculty of Cognitive Sciences and Human Development, University Malaysia Sarawak (UNIMAS). Currently, he is the Director of Centre of Excellence for Semantic Technology and Augmented Reality (CoESTAR) in UNIMAS. His expertise is in the area of Augmented Reality, Virtual Reality and Intelligent Interactive Systems. He has a research interest in the Usability Engineering: Human Computer Interaction, Human Factors in Information Systems, and the Use of Visual Science and Artificial Intelligence to enhance trainer capability, and is overall interested in the social, organizational and technology implications of ICT deployment in organizations. He has more than 10 years experience in running multiple projects and consultations for Teaching, Research, Development and Commercialization in the universities or industries. His research team has invented ARUT Version 1.0, ARUT Mobile Version 1.0 and ARUT Web Version 1.0 for Augmented Reality Applications with international CMMI software development standard. Currently, he is involved in a number of research projects, such as Intelligent Augmented Reality application, Finger Tracking as an Input Devices; Knowledge-Based Augmented Reality for Education, Training, Museum, Planetarium, Tourism, Communication, Games and etc. He has won many national and international awards from his research projects as well as has published novel results intensively in local

and international conferences, seminars, and journals in his field of interest.



Behrang Parhizkar obtained his bachelor degree in Multimedia Technology Management from Multimedia University, Malaysia, followed by master degree of Information Technology & Multimedia System, faculty of Information science and technology, University of Kebangsaan Malaysia (UKM), Bangi, Malaysia. He is a member of Visual Informatics research group. He is a principal lecturer and director of computer research center in Limkokwing University of Creative technology, Malaysia. He won some International awards such as Itex Bronz Medal of Innovation, Malaysia, Gold Medal of Invention, South Korea, and World Intellectual Property Organization (WIPO) Award for Voice Recognition Browser and Marking System. He is interested to research in Augmented Reality, Mobile Technology, Artificial Intelligent, Voice Recognition and Image Processing. He also works on Augmented Reality Technology, as a research assistant.



Arash Habibi Lashakri obtained his bachelor degree in software engineering from Azad University of Lahijan, Gilan, Iran, followed by master degree of computer system & technology, faculty of computer science and information technology, University of Malaya (UM), Kuala Lumpur, Malaysia. Now, he is a PHD candidate of University Technology Malaysia (UTM) in Computer Science and Information Security. He is a reviewer of a few International journal of computer science and information security and Technical Committee member of some international conferences based on computer science and information Security. He is also a member of some funded research projects and he has published more than 30 papers in various international conferences and journals and three books in computer Science. He is interested to research in network security, cryptography, Augmented Reality, and graphical user authentication (GUA) subjects. He also works on wireless security, mobile communication, and optical network bandwidth allocation systems as a research assistant. He is advisor of Limkokwing University of Creative technology computer and mobile research Labs.

Cancelable Biometrics – A Survey

Indira Chakravarthy

Associate Professor, Dept of Computer Science & Engg
Geethanjali College of Engg & Technology
Hyderabad, India
indira_sowmya@yahoo.co.in

Dr.VVSSS.Balaram

Prof & Head, Dept of Information Technology
SreeNidhi Institute of Science & Technology
Hyderabad, India
vbalaram@sreenidhi.edu.in

Dr.B.Eswara Reddy

Associate Professor & Head, Dept of CSE
Jawaharlal Nehru Technological University, Anantapur, India
eswarcejntua@gmail.com

Abstract— In recent times Biometrics has emerged as a reliable, convenient and effective method of user authentication. However, with the increasing use of biometrics in several diverse applications, concerns about the privacy and security of biometric data contained in the database systems has increased. It is therefore imperative that Biometric systems instill confidence in the general public, by demonstrating that, these systems are robust, have low error rates and are tamper proof. In this context, Biometric template security and revocability becomes an important issue. Protecting a biometric template assumes extreme importance because unlike a password or token, which when compromised can easily be revoked or replaced, a biometric cannot be replaced, once it is compromised. Besides if the same biometric trait is used in multiple applications, a user can be potentially tracked from one application to the other by cross matching biometric databases. Cancelable biometrics attempts to solve this problem by constructing revocable biometric templates. This paper attempts to bring out the various methods followed by different researchers towards building such technology.

Keywords- Cancelable biometrics, biometric template, Salting, Biophasing, Noninvertible transforms, Key binding, Key generation.

I. Introduction

Any Biometric, must in general fulfill the criteria of uniqueness, universality, acceptability, collectability and permanence. Permanence is a key feature for biometrics which means a biometric must retain its features in particular the uniqueness, unchanged or acceptably changed, over the lifetime of the individual. However, this very feature of permanence has brought biometrics to challenge a new risk. Conventional authentication methods like passwords and tokens have one great advantage that biometrics do not have viz., they can be cancelled and replaced by a newer version, if ever they were lost or stolen. On the other hand if biometric data is ever compromised from a database, by unauthorized persons, the genuine owner will lose control over it forever and lose his/her identity[1]. This makes the biometric

templates stored in the database stand out as a vulnerability of the authentication system.

A) A successful attack on the biometric template in the database can lead to the following risks:

- i) Template can be replaced by an imposter's template to gain unauthorized access.
- ii) A physical fake can be created from the template to gain access to the system as well as other systems which use the same biometric trait.
- iii) Stolen template can be replayed to the matcher to gain unauthorized access.[6]

B) Therefore the design of a biometric database must be such that, it protects the biometric templates against the above vulnerabilities. Such a Biometric template protection scheme should have the following four properties[9].

- i) **Diversity** : The secure template must not allow cross matching across different databases. This property ensures privacy of user's data.
- ii) **Revocability** : It should be easy to revoke a compromised template and reissue a new template in its place using the same biometric. This property ensures cancelability.
- iii) **Security** : It must be computationally impractical to obtain original biometric template from the secure template. This property ensures that physical spoofing of the biometric is not possible from the stolen template.
- iv) **Performance** : Using the secure template in place of original, should not degrade the performance of the system.
- v) **Intra user variability** : The secure template should accommodate the intra user variability while acquiring and matching the biometric templates during authentication process.

II. Template protection methods

Template protection schemes can be divided into following two categories[5] :

- a) Feature transformation approach – which can be further divided into
 - i) Biometric Salting
 - ii) Non invertible transforms
- b) Biometric cryptosystems – further categorized as
 - i) Key binding
 - ii) Key generation

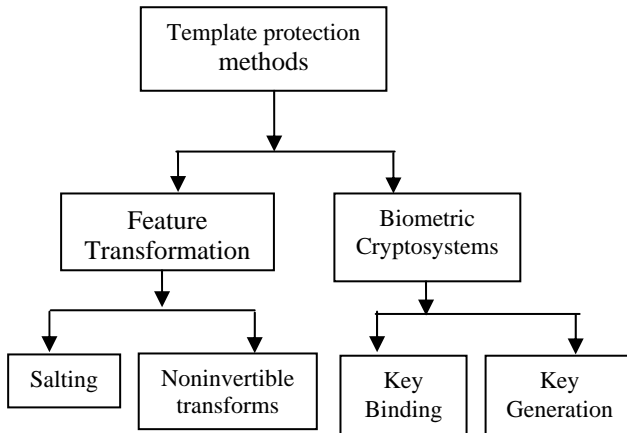


Fig 1. Different approaches to template protection
Following sections detail the design of such schemes.

III. Feature transformation schemes

In this method a transformation function is applied to the original biometric and the transformed template is stored in the system's database instead of the original template. The parameters of the transformation function are typically derived from a random key supplied by the user[6]. Thus the transformed template is represented as $F(T, K)$ where F represents the transformation function T represents the original biometric template and K represents the user supplied parameter.

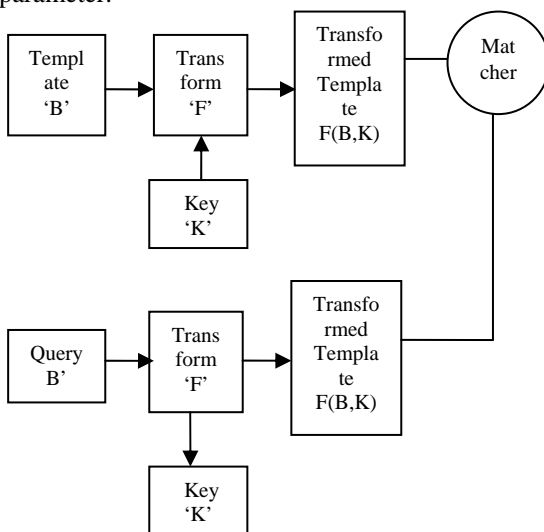


Fig 2 : Matching process with biometric transform scheme

Depending upon the characteristics of the transformation function F , the feature transformation schemes can be further categorized as follows :

A) Biometric Salting : Biometric salting or **Biohashing** is similar to password salting in cryptography. In cryptographic salting the password ' P ' of the user is concatenated with a pseudorandom string ' S ', a hash is taken over the result, and the resulting hash $H(P+S)$ is stored in the database. In Biometric salting, an auxiliary data like a password or user-specific random number is combined with biometric data, a transformation function is applied to this, to derive a transformed version of the biometric template. Since the auxiliary data is externally derived, and is user specific, if the template is ever compromised it can be easily changed and revoked by simply changing the auxiliary data. Additionally since the templates can be different for different applications, if the template is compromised in one application it does not affect the security of other applications.

On the other hand, since the auxiliary information is user specified, user has to remember this and present it at the time of authentication. Hence the security of the salting scheme is based on the secrecy of the key or password. Further the transformation function is not non invertible meaning if an attacker gains access to the key and the transformed template he/she can recover the original biometric template[5]. Teoh et al (2003) proposed a novel two factor authenticator based on iterated inner products between tokenized pseudo-random number and the user specific fingerprint feature, which generated from the integrated wavelet and Fourier-Mellin transform, and hence produced a set of user specific compact code that was coined as BioHashing. BioHashing was shown to be highly tolerant of data capture offsets, with same user fingerprint data resulting in highly correlated bitstrings. Moreover, there was no deterministic way to get the user specific code without having both token, with random data and user fingerprint feature[35].

Savvides et al (2004) proposed a scheme that encrypts the training images used to synthesize the single minimum average correlation energy filter for biometric authentication for face recognition. Different templates can be obtained from the same biometric by varying the convolution kernels thus enabling the cancelability of the templates. They showed theoretically that convolving the training images with any random convolution kernel prior to building the biometric filter does not change the resulting correlation output peak-to-sidelobe ratios, thus preserving the authentication performance. However, the security could be jeopardized via a deterministic deconvolution with a known random kernel[10].

An enhancement of cancelable correlation filter encryption was reported by Hirata and Takahashi (2009). It was shown that the security is heightened by applying Number Theoretic Transform, a Fourier-like transform over a finite field, into biometric data before random kernel convolution[3].

Teoh et al (2004,2006) proposed the random multi-space quantization technique. Their technique extracts the

most discriminative projections of the face template using Fisher discriminate analysis and then projects the obtained vectors on a randomly selected set of orthogonal directions[11]. This random projection defines the salting mechanism for the scheme. To account for intra-user variations, the feature vector obtained after random projection is binarized. The threshold for binarization is selected based on the criteria that the expected number of zeros in the template is equal to the expected number of ones so as to maximize the entropy of the template. The security in this scheme is provided by the user-specific random projection matrix. If an adversary gains access to this matrix, she can obtain a ones so as to maximize the entropy of the template. The security in this scheme is provided by the user-specific random projection matrix. If an adversary gains access to this matrix, she can obtain a coarse estimate of the biometric template [6].

A variant of BioHashing, known as Multistage Random Projection (MRP) (Teoh and Chong, 2007) was proposed to address the stolen-token performance issue. Both theoretical and experimental analysis showed that the performance regresses to the original system under stolen-token scenario[3].

Lumini et al. (2007) improved the performance of BioHashing under stolen-token scenario by utilizing different threshold values and fuse the scores. Their approach improve the base BioHashing in order to maintain a very low equal error rate when nobody steals the Hash key, and to reach good performance even when an “impostor” steals the Hash key.

Lu Leng et al (2005) proposed cancelable PalmCode generated from randomized Gabor filters for palmprint template protection[37].

Jeong et al (2006) proposed a biometric salting scheme for face recognition using an appearance based approach. In their method, an ICA (Independent Component Analysis) coefficient vector is extracted from an input face image. Some components of this vector are replaced randomly from a Gaussian distribution which reflects the original mean and variance of the components. Then, the vector, with its components replaced, has its elements scrambled randomly. A new transformed face coefficient vector is generated by choosing the minimum or maximum component of multiple (two or more) differing cases of such transformed coefficient vectors. If this was compromised, a new feature vector can be generated by changing the permutation matrix.

Lee et al, (2010) proposed a new method to generate cancelable bit-strings from fingerprint minutiae. Their method provides a simple means to generate cancelable templates without requiring pre-alignment of fingerprints. The main idea is to map the minutiae into a predefined 3 dimensional array which consist of small cells and find out which cells includes minutiae. One of minutiae is chosen as a reference minutia and other minutiae are translated and rotated in order to map the minutiae into the cells based on the position and orientation of the reference minutia. The cells in the 3D array are set to 1 if they include more than one minutia otherwise the cells are set to 0. A 1D bit-string is generated by

sequentially visiting the cells in the 3D array. The order of the 1D bit-string is permuted according to the type of reference minutiae and user's PIN so that new templates can be regenerate when needed. Finally, cancelable bit-strings are generated by changing the reference minutia into another minutia in turn.[13]

However the accuracy and vulnerabilities of existing biometric salting schemes needs further justification (Kong et al., 2008).

B) Non-invertible Transforms : In this scheme, a oneway, non invertible function is applied to the original biometric to obtain transformed biometric template. The transformation occurs in the same signal or feature space as the original biometric. The transformation function is so designed that, it is easy to compute in polynomial time but difficult to invert. The parameters of the transformation function are defined by a key which must be available at the time of authentication to transform the query feature set. Since the function is non invertible, even if this key is compromised, it is computationally impossible to invert the transformed template and arrive at the original biometric template. The transformation functions can be application as well as user specific making the biometric highly revocable.

However the main drawback of this approach is the trade-off between discriminability and noninvertibility of the transformation function. Discriminability means, that the transformation function should be such that, features from the same user should have high similarity in the transformed space and features from different users should be quite dissimilar after transformation. Non invertibility feature ensures that it is difficult to obtain the original template from the transformed template. It is difficult to design transformation functions that satisfy both discriminability and non-invertibility conditions simultaneously. Also, the transformation function depends on the biometric features to be used in a specific application.

The invertibility issue, was addressed with **BioPhasoring** (Teoh et al., 2006, 2007). BioPhasor is a set of binary code based on iterated mixing between the user-specific tokenised pseudo-random number and the biometric feature. This method enables straightforward revocation of biometric template via token replacement. The transformation is non-invertible and the BioPhasor is able to achieve extremely low error rate compare to original biometrics in verification setting. The privacy invasion and non-revocable problems in biometrics could be resolved by revocation of resulting feature through the pseudo-random number replacement[13]. Nanni and Lumini (2008) presented a quantized underdetermined non-linear equation system as well as resampled and concatenation of long BioHash with random subspace technique. Other proposals that stem from the idea of user-specific random projection include random correlator (Chong et al., 2006), multiple high dimension random projection (Kim and Toh, 2007), shifted Random Orthonormal Transformation (Wang and Plataniotis, 2007), one-time face template (Lee et al., 2007), 2ⁿ Discretization (Teoh et al., 2008), Preserving

Transform with distinguishing points (Feng et al., 2008b), Sorted Index Numbers (Wang, YJ and Hatzinakos, D., 2009), augmented random projection (Sohn et al. 2009) and a combination of BioHashing and BioPhasor (Nanni and Lumini, 2010) [3].

The defining work in the field of cancelable biometrics was done by Ratha et al (2007). They demonstrated three different methods to generate cancelable fingerprint templates viz., Cartesian, polar, and surface folding transformations. In Cartesian transformation, the minutiae space is divided into a rectangular grid and each cell (possibly containing some minutiae) is shifted to a new position in the grid corresponding to the translations set by the key. The polar transformation is similar to cartesian transformation with the difference that the image is now tessellated into a number of shells and each shell is divided into sectors. Since the size of sectors can be different (sectors near the center are smaller than the ones far from the center), restrictions are placed on the translation vector generated from the key so that the radial distance of the transformed sector is not very different than the radial distance of the original position. For the functional transformation, Ratha et al.[5] used a mixture of 2D Gaussians and electric potential field in a 2D random charge distribution as a means to translate the minutiae points. The magnitude of these functions at the point corresponding to a minutia is used as a measure of the magnitude of the translation and the gradient of a function is used to estimate the direction of translation of the minutiae. In all the three transforms, two or more minutiae can possibly map to the same point in the transformed domain. For example, in the Cartesian transformation, two or more cells can be mapped onto a single cell so that even if an adversary knows the key and hence the transformation between cells, he cannot determine the original cell to which a minutia belongs, because each minutiae can independently belong to one of the possible cells. Also since the transformations used are locally smooth, the error rates are not affected significantly and the discriminability of minutiae is preserved to a large extent.

Based on their empirical results and a theoretical analysis they concluded that feature-level cancelable biometric construction is practicable in large biometric deployments.

Farooq et al. (2007) presented a method by converting the fingerprint minutiae into a cancelable bitstring, without registration or pre-alignment. The idea is based on the fact that fingerprints can be represented by a set of triangles derived from sets of three minutiae that can be used directly in template-based matching. The proposed method is proven to be computational irreversible and satisfies the criteria of reusability and diversity. The reusability is achieved by assigning a unique key to each user in the database to randomize the user template, and in the event of being compromised, the biometric template can be revoked by simply assigning a different key [3].

IV. Biometric Cryptosystems

Biometric cryptosystems, were originally developed for the purpose of either securing a cryptographic key using biometric features or directly generating a cryptographic key from biometric features. However, they can also be used as a template protection mechanism as described in the following sections[34].

A) Key Binding :

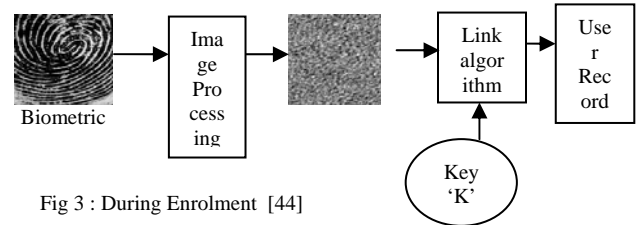


Fig 3 : During Enrolment [44]

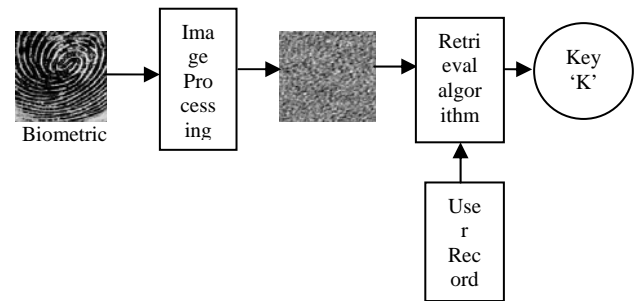


Fig 4 : During Verification [44]

A) Key Binding :

In biometric keybinding schemes, the biometric template is secured by monolithically binding it with a key within a cryptographic framework [6]. During enrolment, the biometric key binding algorithm links a digital key with the biometric to create a secure template known as *User Record*. When the key is required, user presents biometric image to a capture device. The biometric key binding algorithm combines the presented biometric information with user record to retrieve the digital key. Correct key retrieval indicates a match.

Fuzzy commitment scheme [14] proposed by Juels and Wattenberg is a well known example of the key binding approach. Juels and Wattenberg combined techniques from the areas of error-correcting codes and cryptography to achieve a new type of cryptographic primitive which they called **Fuzzy commitment scheme**. The fuzzy commitment scheme is both concealing and binding in that it is not feasible for an attacker to learn the committed value, and also for the committer to decommit a value, in more than one way. In a conventional scheme, a commitment must be opened using a unique witness, which acts, essentially, as a decryption key. By contrast, the fuzzy commitment scheme accepts a witness that is close to the original encrypting witness in a suitable metric,

but not necessarily identical. This characteristic makes the scheme useful for applications such as biometric authentication systems, in which data is subject to random noise. Because the scheme is tolerant of error, it is capable of protecting biometric data just as conventional cryptographic techniques, like hash functions, are used to protect alphanumeric passwords[14]. This scheme has been implemented by Bioscrypt Inc, and is used by their vendors Authentec Inc as sensor provider to Targus, Acer and Synaptics [44].

Equally popular is the concept of **Fuzzy Vault**, introduced by Juels and Sudan [15]. Fuzzy vaults account for two deficiencies in the fuzzy commitment scheme: intolerance of substantial symbol reordering, and security over non-uniform distributions. Briefly explained, Alice places a secret key **K** in a fuzzy vault and locks it using a set **A** of elements from some public universe **U**. To unlock the vault, and retrieve **K**, Bob must present a set **B** that substantially overlaps with **A**. Fuzzy vaults are order invariant, meaning **A** and **B** may be arranged in any order. To protect **K**, it is represented as a polynomial **p**, specifically encoded in the coefficients. A set of points **R** is constructed from **A** and **p(A)**. In addition to these points, chaff points **C** are randomly generated and inserted into **R**. Juels and Sudan solved the subset matching problem with Reed-Solomon coding. To decode **K**, if Bob's **B** approximately matches **A**, he can isolate enough points in **R** that lie on **p** so that applying the error correcting code he can reconstruct **p**, and hence **K** [15]. This scheme has been implemented for fingerprint[16], face[17], iris[18] and signature[19] biometrics.

B) Key Generating Biometric Cryptosystems :

Direct cryptographic key generation from Biometric data is extremely challenging as, such a key cannot be reproduced exactly at the time of verification. This is due to the noise which is inevitably introduced during biometric sample acquisition.

The defining feature of key generating biometric cryptosystems is the use of two functions called Generating and Reproducing functions. Broadly speaking, the generating function takes the biometric data along with user specific key/information 'K' to produce a public string 'P' and a secret string 'S', $Gen(B, K) \rightarrow \langle S, P \rangle$. The reproducing function takes the public string along with query biometric measurement to reproduce the secret string ie $Rep(B', P) \rightarrow S$. In other words, the scheme extracts some randomness 'S' from 'B' and then successfully reproduce 'S' as long as $d(B, B') \leq \epsilon$, where 'd' is a metric $d(B, B')$ (e.g., Hamming distance, Euclidian distance, set distance etc.,) on noisy biometric data B and query biometric B'.

Dodis et al [21] coined the terms **Secure Sketch** and **Fuzzy extractor** in the context of key generation from biometric data. A secure sketch is helper data that gives only limited information about the template even in the worst case (i.e., the entropy loss should be low) but allows exact reconstruction of the template when a biometric query close to the stored template is presented. Fuzzy extractor on the

otherhand is a cryptographic primitive that generates a cryptographic key from the biometric features.

Dodis et al proposed constructions and rigorous analysis of secure sketches for three metrics viz., Hamming distance, Set difference and Edit distance. Qiming Li, Yagiz Sutcu, and Nasir Memon studied the entropy loss due to quantization. This occurs when a biometric template, represented as points in continuous domains with unknown distributions, is quantized (discretized) and a known sketch scheme is applied in the discrete domain. They analyzed the entropy loss due to quantization and tried to arrive at the "optimal" quantizer[24].

Chang and Li in their studies [25], considered two aspects namely a) quantization and b) the issues of authentication, forgery and preimage attacks. To handle the first issue, they considered using two levels of quantization. The second issue leads to the proposed additional requirement on *sensitivity*. Their study concentrated on how to choose the optimal parameters under the trade-off of robustness, size and sensitivity, and show that in many practical settings, the two-level quantization can be significantly more effective than a seemingly natural method of assigning one bit to each coefficient. Buhan et al addressed the problem of generating fuzzy extractors by modeling the biometric data more naturally as a continuous distribution [26]. Their study showed that there is a direct relation between the maximum length of the keys extracted from biometric data and the error rates of the biometric system. The length of the bio-key depends on the amount of information that can be extracted from the source data. This information can be used a-priori to evaluate the potential of the biometric data in the context of a specific cryptographic application.

There have been a number of works on how to extract consistent keys from handwritten online signatures [26], fingerprints [30], iris patterns [27], voice features [28], and face biometrics [29], multimodal systems (face and fingerprint) [31]. These, however, do not have sufficiently rigorous treatment of the security, compared to well-established cryptographic techniques. Some of the works give analysis on the entropy of the biometrics, and approximated amount of efforts required by a brute-force attacker.

V) Summary of different schemes to secure biometric template

Table 1 below summarises the different schemes to secure biometric templates in terms of Template protection method, key principle, public domain used, advantages and disadvantages in each of these methods.

Table 1 : Comparison of various biometric template protection schemes

Template protection Method	Key Principle	Public Domain	Advantages	Disadvantages
Salting	Secrecy of key 'K'	Transformed template $F(T;K)$	a) Since Key is user specific, multiple templates for the same user can be generated. b) If a template is compromised it can be easily revoked and replaced with a new template using a different key.	a) As the transformation is invertible, if the key is ever compromised, template is no longer secure. b) Matching takes place in transformed domain. Therefore the scheme has to be designed in such a way that recognition performance does not degrade ; especially in the presence of large intra user variation.
Non-invertible Transform	Non-invertibility of the transformation function F	Transformed template $F(T;K)$	a) Better security than salting since it is extremely difficult to recover the original biometric template even if the key is compromised. b) Transformation function can be designed to be application specific and / or user specific. This ensures diversity and revocability of biometric templates.	a) The transformation function should be such that features from the same user should have high similarity in the transformed space and features from different users should be quite dissimilar after transformation. In addition given a transformed feature set it should be hard to obtain the original feature set. The tradeoff between discriminability and noninvertibility of the transformation function forms the main drawback of this approach.
Key-binding biometric Cryptosystem	Level of security depends on the amount of information revealed by the helper data 'H'	Helper Data $H = F(T;K)$	a) Tolerant to intra-user variations in biometric data. The tolerance is determined by the error correcting capability of the associated codeword	a) Matching has to be done using error correction schemes and this precludes the use of sophisticated matchers developed specifically for matching the original biometric template. Can possibly lead to a reduction in the matching accuracy. b) In general, biometric cryptosystems are not designed to provide diversity and revocability. However, attempts are being made to introduce these two properties into biometric cryptosystems by using them in conjunction with other approaches such as salting . c) The helper data depends on the specific biometric features to be used and the nature of associated intra-user variations.
Key-generating biometric cryptosystem	Level of security depends on the amount of information revealed by the helper data 'H'	Helper Data $H = F(T;K)$	a) Direct key generation from biometrics is an appealing template protection approach which can also be very useful in cryptographic applications.	a) It is difficult to generate key with high stability and entropy, due to the noise which is inevitably introduced during biometric sample acquisition .

VI. Recent Developments

A) TURBINE (TrUsted Revocable Biometric IdeNtitiEs - 2007) : Turbine is a research project, awarded 6.3 Million Euro funding by the European Union under the Seventh Framework Programme (FP7) for Research and Technology Development. The TURBINE consortium comprises major players in biometrics and cryptography, including Morpho (ex Sagem Sécurité) , Philips Research Europe, Morpho e-Documents, Precise Biometrics in Sweden, Cryptolog and ARTTIC in France, 3D-GAA S.A. in Greece, as well as academic research groups from Katholieke Universiteit Leuven in Belgium and Gjøvik University College in Norway[43].

Originally planned for three years, TURBINE aims to develop innovative digital identity solutions, combining a) secure, automatic user identification and b) reliable protection of the biometrics data through advanced cryptography technology. Research focus is on transformation of a fingerprint, so that the result can only be re-generated by the person with the fingerprints. TURBINE will hence provide assurance that:

- i) The data used for the authentication, generated from the fingerprint, cannot be used to restore the original fingerprint sample
- ii) The individual will be able to create different "pseudo-identities" for different applications with the same fingerprint, whilst ensuring that these different identities cannot be linked to each other, and
- iii) The individual is enabled to revoke an identity for a given application in case it should not be used anymore.

The outcome of the project is intended to meet usage requirements for various market segments, such as ebanking, eGovernment, eHealth, physical access control, and mobile telecommunications.

B) VAST LAB (Vision and Security Technology , Recently field preliminary patent , currently working on spin-off company by Dr . Terry Boulton ,) : Dr Boulton has developed an approach that allow biometric data to be converted to a secure but revocable form that still allows the computation of robust distance needed for effective biometric data. A variation supports identification but cannot be used for recognition, i.e. a fingerprint-based biometric that can prove you are you but cannot be used by anyone to look for you in a database or to link two databases. The result is a technique that preserves privacy but can enhance security[38].

C) HITACHI : On July 24 , 2007 Hitachi announced a biometric cardless credit payment system, called "finger vein money" , which allows shoppers to pay for purchases using only their fingertips. Finger vein money relies on Hitachi's finger vein authentication technology, which verifies a person's identity by reading the pattern of blood vessels in his or her fingers. These blood vessel patterns are unique to each



Fig 6 : The mechanism of TURBINE process[43].

individual, and are hidden securely under the skin, making them all the more difficult to counterfeit. Hitachi's finger vein authentication technology is already being used to verify user identities for ATMs, door access control systems and computer log-in systems in Japan and elsewhere.

In the finger vein money system, consumers first register their finger vein pattern data with the credit card company. The data is then entered into a database along with the individual's credit account information. Later, when shoppers want to pay for something, they simply go to the cash register and place their finger in a vein reader, which uses infrared LEDs and a special camera to capture a detailed image of their vein structure. The image is converted into a readable format and sent to the database, where it is checked against the records on file. When the system verifies the identity of the shopper, the purchase is charged to the individual's credit account. Hitachi is conducting the trial with the cooperation of major credit card company JCB .

To protect the biometric data in this system , Hitachi used encryption algorithm called Correlation Invariant Image Randomization (CIIR) and matches the encrypted data to an encrypted template without decrypting the data. This keeps the biometric data secret from eavesdroppers as well as administrator of server in a remote biometric system. Additionally , even if the stored data is compromised , it can be cancelled and replaced by simply changing the encryption key , resulting in a secure biometric authentication system[39].

D) Priv-ID : priv-ID , originated from Royal Philips Electronics and is based at the High Tech Campus, Eindhoven, the Netherlands, is the leading provider of PET (Privacy Enhancing Technology) that eliminates privacy and security concerns in biometric deployments. The company offers high-quality BioHASH® solution, which stores and matches standardized fingerprint information using an irreversible binary hash code. On December 3, 2010, priv-ID released a biometric Match-on-Card solution based on its BioHASH® technology. This Match-on-Card implementation is based on a fundamentally different approach, leading to an absolute minimum code-size requirement, while providing portability to different cards and a very high matching speed without compromising matching accuracy. priv-ID's match-on-card is based on the successful BioHASH® technology, that transforms the biometrics into a binary feature vector, that can

be protected with an off-the-shelf cryptographic hash function such as SHA-256. The BioHASH® matcher is modality independent and works the same for fingerprint, iris, face or vein information[40].

VII. Conclusion and scope for further research

Biometric technology creates a one-to-one correspondence between a person and a record, thus providing a natural tool for identity management. However widespread deployment of biometrics for a variety of applications has given rise to apprehensions among the public, that biometric technologies may invade privacy. Further since a biometric is a permanent feature, associated with a person, once a biometric template in the database is compromised, it is lost/compromised forever.

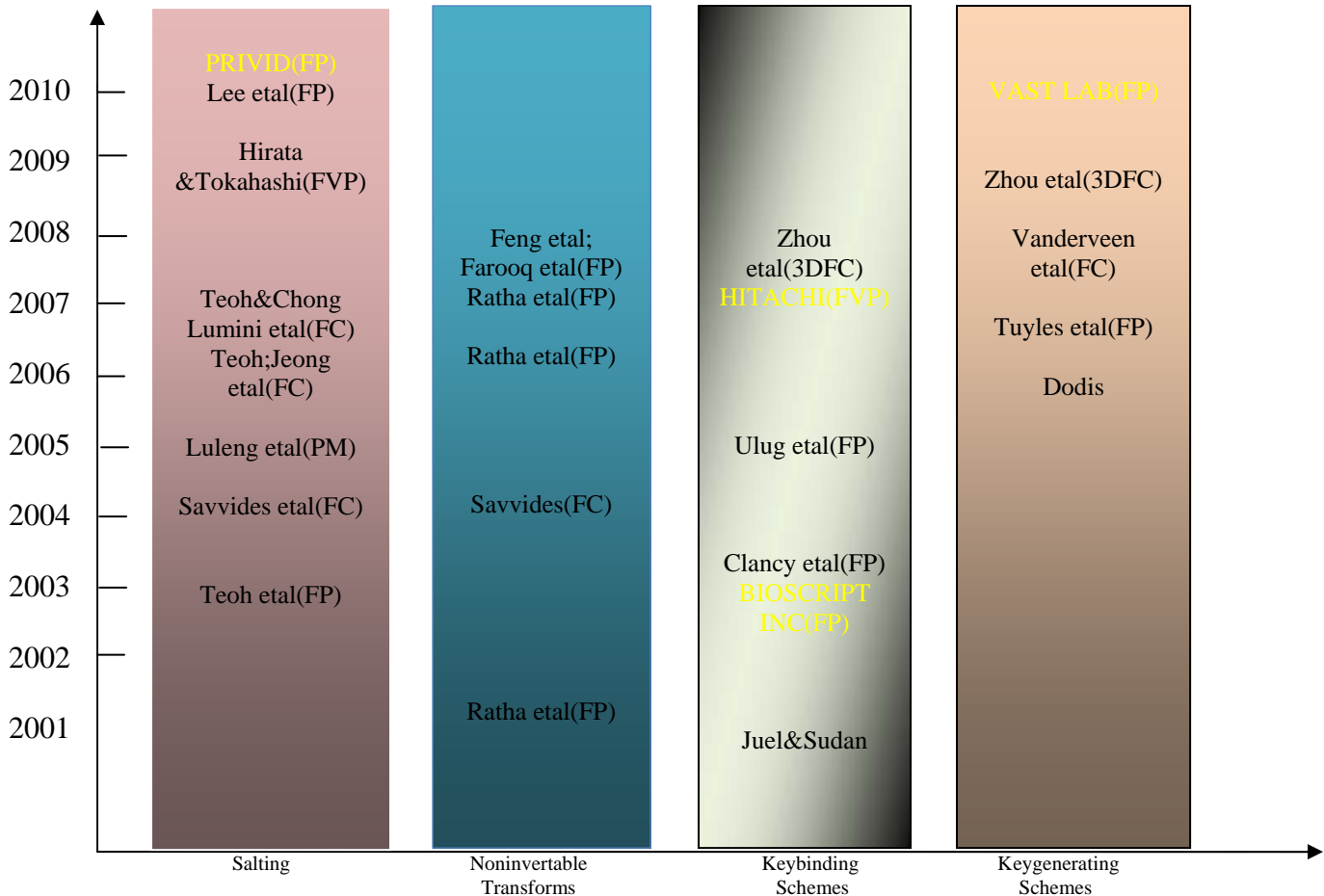
Unless these controversies surrounding biometrics are addressed convincingly, there is a danger that biometric authentication method may lose its popularity with the general public. Hence biometric schemes have to be designed in such a way that they instill confidence in public with respect to the

privacy and security of the template stored. Thus biometric template protection and revocability is a very important complement for biometric systems.

This paper describes various template protection schemes and the revocability nature of such schemes available in literature and discussed their relative advantages and drawbacks. commercial implementations of such schemes is also included wherever possible.

However the available template protection schemes are not yet sufficiently mature for large scale deployment; they do not meet the requirements of diversity, revocability, security and high recognition performance[5]. Also a rigorous analysis of template security schemes, with the exception of Biohashing has not been taken up. such an analysis is a must before the template security scheme can be deployed in critical real world applications. Further scope for research exists in the area of non invertible transforms.

Finally instead of a single template protection approach, a hybrid scheme that makes use of the advantages of the different template protection approaches must be developed.



Black text indicates research ; Yellow text indicates implementations

Fig 7 : Progress of research and Implementations in the past decade .

References

- [1]. Rudd M.Bolle, Jonathan H.Connell,Sharath Pankanti, Nalini K. Ratha, Andrew W. Senior, Guide to biometrics, Springer Publication (2003).
- [2]. Davide Maltoni, Anil K. Jain, Handbook of fingerprint recognition, Springer publication (2002).
- [3]. N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM systems Journal*, vol. 40, pp. 614-634, 2001.
- [4]. Nalini K.Ratha,Sharat Chikkerur,Jonathan H.Connell and Ruud M.Bolle Generating cancelable fingerprint templates. *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, april 2007
- [5]. Biometric Template Security Anil K. Jain, Karthik Nandakumar and Abhishek Nagar, *EURASIP Journal on Advances in Signal Processing*, Special Issue on Biometrics, January 2008
- [6]. A. K. Ross, J. Shah, and A. K. Jain, "From Templates to Images:Reconstructing Fingerprints From Minutiae Points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*,vol. 29, no. 4, pp. 544-560, 2007.
- [7]. A. Adler, "Images can be Regenerated from Quantized Biometric Match Score Data," in *Proceedings Canadian Conference on Electrical and Computer Engineering*, Niagara Falls, Canada, May 2004, pp. 469-472.
- [8]. D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer-Verlag, 2003.
- [9]. Marios Savvides, B. V. K. Vijaya Kumar, P. K. Khosla, "Cancelable Biometric Filters for Face Recognition," *Pattern Recognition, International Conference on*, vol. 3, pp. 922-925, 17th International Conference on Pattern Recognition (ICPR'04) - Volume 3, 2004.
- [10]. A. B. J. Teoh, A. Goh, and D. C. L. Ngo, "Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity inputs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.28,no. 12, pp. 1892-1901, December 2006.
- [11]. CANCELLABLE FACE BIOMETRICS SYSTEM by COMBINING INDEPENDENT COMPONENT ANALYSIS COEFFICIENTS, MINYI JEONG AND ANDREW BENG JIN TEOH, LECTURE NOTES IN COMPUTER SCIENCE, 2011, VOL 6540.
- [12]. Chulhan Lee, Jaihie Kim Cancelable fingerprint templates using minutiae-based bit-strings, *Journal of Network and Computer Applications archive* Volume 33 Issue 3, May 2010
- [13]. Biophasor: Token Supplemented Cancellable Biometrics. Teoh, A.B.J., Ngo, D.C.L., *Control, Automation, Robotics and Vision*, 2006. ICARCV '06. 9th International Conference on 5-8 Dec 2006.
- [14]. Ari Juels, Martin Wattenberg; A fuzzy commitment scheme, *Proceeding Conference on computer and communications security*, ACM NewYork
- [15]. A. Juels and M. Sudan, "A Fuzzy Vault Scheme," in *Proc. IEEE Int. Symp. On Information Theory*, 2002, p.408.
- [16]. S. Yang and I. Verbauwhede, "Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, Philadelphia, March 2005, pp. 609-612.
- [17]. Y. C. Feng and P. C. Yuen, "Protecting Face Biometric Data on Smartcard With Reed-Solomon Code," in *Proceedings of CVPR Workshop on Biometrics*, New York, USA, June 2006, p. 29.
- [18]. Y. J. Lee, K. Bae, S. J. Lee, K. R. Park, and J. Kim, "Biometric Key Binding: Fuzzy Vault based on Iris Images," in *Proceedings of Second International Conference on Biometrics*, Seoul, South Korea, August 2007, pp. 800-808.
- [19]. M. Freire-Santos, J. Fierrez-Aguilar, and J. Ortega-Garcia, "Cryptographic Key Generation Using Handwritten Signature," in *Proceedings of SPIE Conference on Biometric Technologies for Human Identification*, vol. 6202, Orlando, USA, April 2006, pp. 225-231.
- [20]. Secure Sketch for Biometric Templates Qiming Li, Yagiz Sutcu, and Nasir Memon.
- [21]. Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data," *Cryptology ePrint Archive*, Tech. Rep. 235, February 2006.
- [22]. Fuzzy Extractors_ A Brief Survey of Results from 2004 to 2006 Yevgeniy Dodis, Leonid Reyzin, Adam Smith, April 9, 2008 A prior version of this article appears as Chapter 5 of [25] New York University, <http://cs.nyu.edu/~dodis/>. Supported by the National Science Foundation grants CCR- 0133806 and CCR-0311095, and by the New York University Research Challenge Fund 25-74100-N5237. Boston University, <http://www.cs.bu.edu/~reyzin>. Supported by National Science Foundation grants CCF- 0515100 and CNS-0546614, and the Institute for Pure and Applied Mathematics at UCLA. Pennsylvania State University, <http://www.cse.psu.edu/~asmith>. This survey was written while L.R. and A.S. were visiting the Institute for Pure and Applied Mathematics at UCLA.
- [23]. P. Tuyls, B. Skoric, and T. Kevenaar, editors. *Security with Noisy Data*. Springer-Verlag, 2007.
- [24]. Qiming Li, Yagiz Sutcu, and Nasir Memon qiming.li@ieee.org ygzstc@yahoo.com
- [25]. Qiming Li Ee-Chien Chang, Robust, short and sensitive authentication tags using secure sketch *Proceeding, MM&Sec '06 Proceedings of the 8th workshop on Multimedia and security ACM*
- [26]. F. HAO AND C.W. CHAN. PRIVATE KEY GENERATION FROM ON-LINE HAND WRITTEN SIGNATURES. *INFORMATION MANAGEMENT AND COMPUTER SECURITY*, 10(2), 2002.
- [27]. Feng Hao, Ross Anderson, and John Daugman. Combining cryptography with biometrics effectively. Technical Report UCAM-CL-TR-640, University of Cambridge, 2005.
- [28]. F. Monrose, M.K. Reiter, Q. Li, and S. Wetzel. Cryptographic key generation From voice. In *IEEE Symp. on Security and Privacy*, 2001.
- [29]. Y. Sutcu, T. Sencar, and N. Memon. A secure biometric authentication scheme Based on robust hashing. In *ACM MM-SEC Workshop*, 2005.
- [30]. Shenglin Yang and Ingrid Verbauwhede. Automatic secure fingerprint verification system based on fuzzy vault scheme. In *IEEE Intl. Conf. on acoustics, Speech, and Signal Processing (ICASSP)*, pages 609-612, 2005.
- [31]. Y. Sutcu, Q. Li, and N. Memon, "Secure Biometric Templates from Fingerprint-Face Features," in *Proceedings of CVPR Workshop on Biometrics*, Minneapolis, USA, June 2007.
- [32]. U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric Cryptosystems: Issues And Challenges," vol. 92, no. 6, June 2004.
- [33]. Andrew Teoh Beng Jin, David Ngo Chek Ling, Alwyn Goh; Biohashing: two factor authentication featuring fingerprint data and tokenised random number, 2004 *Pattern Recognition Society*. Published by Elsevier Ltd.
- [34]. *ENCYCLOPEDIA OF BIOMETRICS, VOLUME 2*, BY STAN Z. LI, ANIL K. JAIN, LIBRARY OF CONGRESS NO :2009929415; SPRINGER SCIENCE + BUSINESS MEDIA, LLC, 2009
- [35]. Lu Leng, Jiashu Zhang, Muhammad Khurram Khan, Xi Chen, Ming Ji and Khaled Alghathbar, Cancelable PalmCode generated from randomized Gabor filters for Palmprint Template protection, *Scientific Research and Essays* Vol. 6(4), pp. 784-792, 18 February, 2011 Available online at <http://www.academicjournals.org/SRE> ISSN 1992-2248 ©2011 Academic Journals
- [36]. Shinji Hirata, Kenta Takahashi, Cancelable biometrics with perfect secrecy for correlation based matching, www.arnetminer.org
- [37]. Lecture notes by Alexander Nouak Head Of Department Security Technology Fraunhofer Institut graphische datenverarbeitung, Germany.
- [38]. www.vast.uccs.edu/biodistmet.html
- [39]. www.hitachi.com/technology_rd.pdf
- [40]. http://www.priv-id.com/priv-id-introduces_martcards/
- [41]. Rahul Chaurasia, Symbiosis Law School, Pune, Identity Theft Knowledge Base blog
- [42]. www.scholorpedia.com
- [43]. www.turbine-project.eu/
- [44]. www.biometrics.org/bc2004/CD/PDF/PROCEEDINGS/

AUTHORS PROFILE



Indira Chakravarthy graduated from Osmania University college of Engg, Hyderabad and Post graduated in Computer Science & Engg from Osmania University college of Engg, Hyderabad in 1998. She worked as software Engineer thereafter and later shifted to academics in 2004. She is currently attached with Geethanjali college of Engineering and Technology in Hyderabad as Associate

Professor in CSE department. Her areas of interest include Biometrics, Information Security, Database Management Systems

and Software Engineering. Presently she is pursuing Ph.D from Jawaharlal Nehru Technological University, Anantapur, India , in the field of Biometrics.



Dr. V V S S S Balam is currently with Sreenidhi Institute of Science and Technology, Hyderabad , India , working as Professor and Head in the Department of Information Technology. He has 17 years of teaching experience. He did his M.Tech from Andhra University and Ph.D from Osmania University. His areas of interest

include Network Security and Cryptography, Data warehousing and Mining, Operating Systems, Distributed Operating Systems and Computer Graphics. He has a few International Publications to his credit.



Dr. B. Eswara Reddy Graduated in B.Tech.(CSE) from Sri Krishna Devaraya University in 1995. He received Masters Degree in M.Tech.(Software Engineering), from JNT University, Hyderabad, in 1999. He received Ph.D in Computer Science & Engineering from JNT University, Hyderabad, in 2008. He served as Assistant Professor from 1996 to 2006. He is working as Associate Professor in CSE Dept., since 2006 and currently acting

as Head of CSE Dept. at JNTUACE, Anantapur. He has more than 10 Publications in various International Journals and 15 Publications in various National and International Conferences. He is one of the author's of the text book titled Programming with Java published by Pearson/Sanguine Publishers. His research interests include Pattern Recognition & Image Analysis, Data Warehousing & Mining and Software Engineering. He is a life member of ISTE, IE, ISCA and member of CSI and IEEE

Reusable Code for CSOA-Services: Handling Data Coupling and Content Coupling

Shakeel Ahmad¹, Sheikh Muhammad Saqib¹, Muhammad Ahmad Jan¹, Muhammad Zubair Asghar¹ and Bashir Ahmad¹

¹Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

Abstract— ‘Customized-SOA services are loosely coupled’; this should not be a slogan but be physically present. For adopting such features, code for each service should be independent. Common coupling, external coupling, control coupling can be eliminated but data coupling and content coupling may not be removed. But these two coupling types can be handled in such a way that code for each CSOA-service can be treated as a single unit. Here author proposed a strategy to design a code for Customized-SOA service with handling data coupling and content coupling. Such type of code for a CSOA-service will be very useful for reusing service in another CSOA-application. And also such code can be easily changed according to needs of users.

Keywords- Customized-SOA, Loosely Coupling, Data Coupling, Content Coupling, Reusability.

I. INTRODUCTION

Loosely coupling task is not easier, without familiar to understanding the principles it can not be achieved at first step because code is ‘hard to trace’. For writing code, depend on only abstraction not on concrete classes then combined with dependency injection principle, true loosely coupling can be achieved [1]. Interpreted languages often include dynamic execution model and support for functions as first-class values and closures, these features can be very useful to build different coordination mechanisms for distributed asynchronous computing and loosely coupled interactions often supplied by event-oriented application models [3].

SOA model plays a vital role in software design & architecture and through dynamic discovery, it shares different components. There are different methods which enables complex application to be combined with services. Selection of these services depends on cost and quality. So there will be a mechanism which can easily select the services for an application [6]. If services are loosely coupled then their cost will be minimum and quality will be maximum.

Customized-SOA is a services based application i.e. software which has only logical interaction with user and there is no interface for physical entry” is called CSOA Customized SOA application. It provides only services to an organization according to their needs. Business Oriented Services Model for CSOA made up of SOMA and RUP which is Consist of four parts:

BOSI (Business Oriented Service Identification): BOSI presents service identification for service decomposition and

provides an approach for finding completed services before development and reusing these services in other CSOA.BOSS (Business Oriented Service Specification): Goal of BOSS is to specify those services which should be implemented during particular iteration functional Components & technical components. BOSR (Business Oriented Service Realization): BOSR includes managing resources, controlling operations to optimize costs, schedules, and quality. BOSD (Business Oriented Service Deployment): This includes the deployment of the services into the productive environment and user acceptance tests [4]. Since services are loosely coupled in nature, but there may be some problems which can be occurred when service of one application is going to be incorporated in other application. Because some classes uses the features of some other classes. But when services are created at design time with all future references or parameters then these problems can be reduced.

In BOSM, BOSI also discover either developed service can be used in single organization or other organizations with percentage. Here BOSS is very important because functional and technical components contain different classes with their attributes and events. Here author proposed a technique how to implement a service for COSA with its loosely coupled features. So it can be easily reused in another CSOA application.

Different types of coupling are in practice such as common coupling (common coupling exists when classes in the system share a global data), external coupling (dependency of the class on third party classes), control coupling (control coupling is when one class controls the logic of another class), content coupling (exists between two classes when one class relies on the internal working of another class, i.e., one class is using the variables of another class), data coupling (one class passes simple data to another as an argument) [2].

Since in customized-SOA service is consumed by the end user and does not require any further processing and also it is self defined. So there will be little bit or no chance of common coupling, external coupling, control coupling. But content coupling and data coupling can not be avoided for designing the classes of CSOA-services. Here author define a strategy that these content coupling and data coupling will not affect the service when it is going to be reused.

II. STRUCTURE FOR SERVICE CODE

Customized-SOA service is just like a separate unit so avoid inheritance, level-of-inheritance and multiple-

inheritance. Now first of all separate the common features of all services, as information about each service must be stored in database, so each service requires connection variable and queries. These queries may be executable (returns records from database) or non-executable (only make changes in database and returns only integer). Hence there may be two classes 1st class (method for establishing connection) relate to database connection and 2nd class (will contain methods for queries) relate to these executable and non executable queries. 3rd class will be for developed service can be extracted from technical and functional components based on “story cards practice of BOSM in BOSI” [4]. Technical and functional components will contain all possible attributes and events which will be used in current and future application. Description of each event will be useful for creating algorithm about service. Then this algorithm will show the shadow of actual code for service. This target class can use the methods of 1st class and 2nd class with their objects. 1st and 2nd classes will be mandatory for each service so we can not avoid content coupling and data coupling. When 3rd class will be reused in another application, then content coupling and data coupling can be easily removed or skipped.

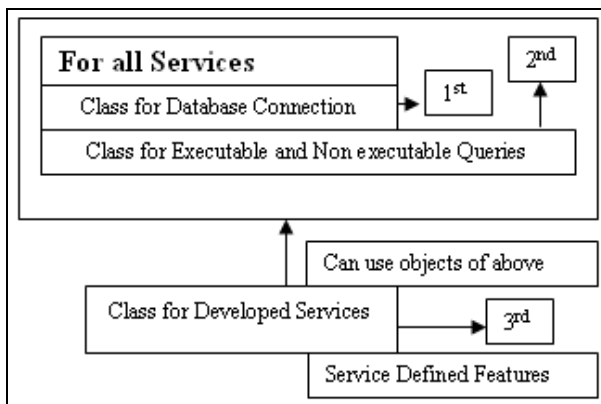


Figure 1. Class structure for CSOA Service.

III. IMPLEMENTATION OF SERVICE CODE

DOA is a CSOA based application developed in .Net framework which can be used in any office and can assist all employees. Each user can perform different activities of DOA after authentication [5]. Attendance is a global service for all organization, now we construct attendance service for DOA, then its code will be loosely coupled for other services.

Table-1: FUNCTIONAL COMPONENT FOR ATTENDANCE SERVICE

Service Component name	Request Messages	Response Messages	Inputs	Inputs Description
clsAttend ance	Attendance Marking	Attendance is at time or Not	intUIDS	Sender ID
			tmAttendance	Attendance Time
			dtAttwndance	Attendance Day

This table contains all possible attributes which can be used in all organization such as any user can make attendance on each day at particular time.

Table-2: TECHNICAL COMPONENTS FOR ATTENDANCE SERVICE.

Service Component name	Actions	Events
Service-1	This event will check the closing time of office and will return false if time is up.	Bool evtCheckTime()
	This event will save the marking attendance of user in database.	evtAttenSetting (intUIDS ,tmAttendance, dtAttwndance)

This table shows the different events related to the attendance service with passing generalized parameters. From above Table-2 class diagram can be easily created.

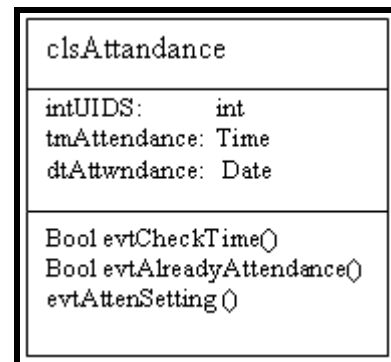


Figure 2. Class Diagram for Service Attendance.

Hence from Fig-1, it is clear that two classes will be mandatory for each new service. Now we are creating 1st, 2nd and 3rd class of Fig-1 in VB.Net framework. Fig-3 is showing the code for connection class. This class contains only connection with database.

```
Public Class dbConnection
    Public cn As System.Data.SqlClient.SqlConnection
    Sub New()
        cn = New System.Data.SqlClient.SqlConnection("integrated
        security=sspi;persist security info=false;initial catalog=dbDOA")
    End Sub
End Class
```

Figure 3. 1st class for Connection with Database.

Each service will have executable and non-executable queries, so 2nd class will be beneficial for creating each new service.

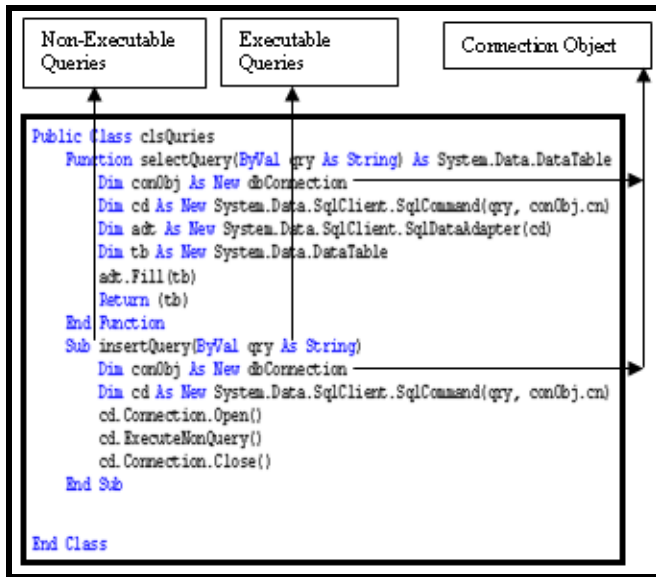


Figure 4. 2nd class for Queries with Database.

Above Fig-4 is using the object of connection class and these two classes will be common among all classes of service. Now from Fig-2, we can easily create a reusable code for service attendance i.e. 3rd class.

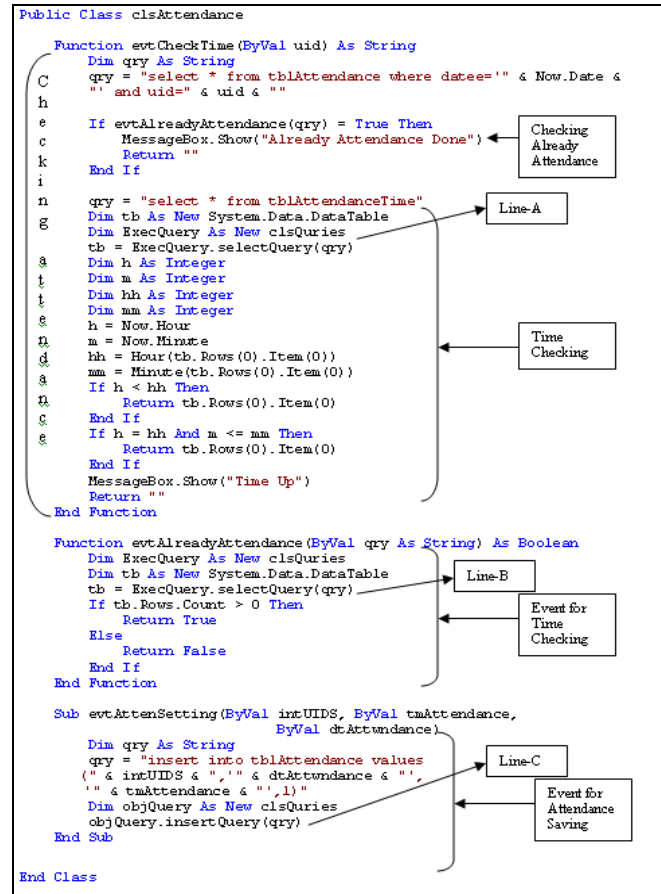


Figure 5. Reusable Code for Service Attendance.

In Fig-5, Line-A, Line-B are executable queries and Line-C is non executable query. In these three lines we are creating just an object of class 'clsQueries' and then calling relevant query event. Database connection depends on class 'clsQueries', how it is establishing the connection with database. Class in Fig-5 must have knowledge about class 'clsQueries' that what will be the internal execution in this class, this is content coupling. Since, we are using object of another class (clsQueries) in clsAttendance class, so here is data coupling. When this service will be plugged in another CSOA type application, then there will be little bit changes required for handling this content coupling and data coupling i.e. Line-A, Line-B and Line-C. Since each application has its own scheme of code for database connections and queries. So this service will be reused only by redesigning lines A, B, C.

IV. CONCLUSION

Some time already developed code can not be used in another application due to some sort of coupling. As major characteristic of SOA and CSOA, services are 'loosely coupled' because services are independent in nature. But at the time of reusing such services in another application, code for each class may be hard to use for incorporating such services in another application. Proposed technique for coding of such services is handling data coupling and content coupling, so those services will be easily reusable or changeable for other applications.

REFERENCES

- [1] Sean Feldman, "Strive for loosely coupled designs between objects that interact", 2008
<http://weblogs.asp.net/sfeldman/archive/2008/06/09/too-loosely-coupled-code.aspx>
- [2] Milind Shingade, "Design Loosely Coupled Modules - Object Oriented Systems", 2007.
http://www.codeproject.com/KB/architecture/Designing_loosely_coupled.aspx
- [3] B. Silvestrea, "Flexibility and coordination in event-based, loosely coupled, distributed systems", Computer Languages, Systems & Structure, 2010
journal homepage: www.elsevier.com/locate/cl
- [4] Sheikh Muhammad Saqib, "MS-Thesis", ICIT, Gomal University PAKISTAN, 2011.
- [5] Shahid Hussain, "Mapping of SOA and RUP: DOA as Case Study", JOURNAL OF COMPUTING, VOLUME 2, ISSUE 1, JANUARY 2010, ISSN 2151-9617, 2010
- [6] Daniel A. Menasc a; "On optimal service selection in Service Oriented Architectures", Performance Evaluation, Volume 67, Issue 8, August 2010, Pages 659-675

About Author

Dr. Shakeel Ahmad received his

B.Sc. with distinction from Gomal University, Pakistan (1986) and M.Sc. (Computer Science) from Qauid-e-Azam University, Pakistan (1990). He served for 10 years as a lecturer in Institute of Computing and Information Technology (ICIT), Gomal University Pakistan. Now he is serving as an Assistant Professor in ICIT, Gomal University Pakistan since 2001. He is among a senior faculty member of ICIT. Mr. Shakeel Ahmad received his PhD degree (2007) in Performance Analysis of Finite Capacity Queue under Complex Buffer Management Scheme. Mr. Shakeel's research has mainly focused on developing cost effective analytical models for measuring the performance of complex queuing networks with finite capacities. His research interest includes Performance modeling, Optimization of congestion control techniques, Software Engineering, Software Refactoring, Network security, Routing Protocols and Electronic learning. He has produced many publications in Journal of international repute and also presented papers in International conferences

Computing the Efficiency of a *DMU* with Stochastic Inputs and Outputs Using Basic DEA Models

M. Nabahat

Sama technical and vocational
training school, Islamic Azad
university, Urmia branch, Urmia, Iran
mehrddad_nabahat@yahoo.com

F.Esmaeeli sangari

Sama technical and vocational
training school, Islamic Azad
university, Urmia branch, Urmia, Iran
fardin_e_s@yahoo.com

S.M.Mansourfar

Sama technical and vocational
training school, Islamic Azad
university, Sahand branch, Sahand,
Iran
mostafa.mansourfar@gmail.com

Abstract- This paper tries to introduce essential models of stochastic and deterministic models (DEA) using the Chance Constrained Programming model to measure the DMU, that enter the system simultaneously with the stochastic inputs and outputs. Considering the fact that by adding a Decision Making Unit to the series of DMUs the efficiency frontier may change, the main goal is to change the available essential stochastic models of DEA and compute the amount of efficiency of the DMU and its position in respect to the old frontier. Finally an Example is shown to highlight the procedure of changing the stochastic model to deterministic model.

Keywords- DEA; Stochastic Programming; Efficiency; frontier

1. Introduction

Stochastic programming is a framework of modeling optimization problems that involve uncertainty. Whereas deterministic optimization problems are formulated with known parameters, real world problems almost invariably include some unknown parameters. Stochastic programming models are similar in style but take advantage of the fact that probability distributions governing the data are known or can be estimated.

This branch of science has been used since the late 1950s for decision models where input data (coefficients in Lp problem) have been given probability distribution. Without any attempt at completeness, we might mention from the early contributions to this field. The pioneer works were done by Dantzing[1-2], Beale[3-4], Tintner[5], Simon, Charnes, Cooper [6-7], Avriel and Williams [8] and Since then ,a number of Stochastic programming models have been formulated in inventory theory, system maintenance ,micro-economics ,banking and finance.

Among various stochastic models, we would like to introduce the chance constrained programming, and we will use it in one of the essential models.

In this paper new DMU with stochastic inputs and outputs is added to the set of observed DMUs then information about efficiency of this DMU and changes in frontier of production possibility set of observed DMUs are interested.

2. Preliminaries

Consider a set of homogenous *DMUs* as DMU_j ($j=1,2,\dots,n$). Each *DMU* Consumes m inputs to produce S outputs. Suppose that $X_j = (x_{1j}, \dots, x_{mj})^T$ and $Y_j = (y_{1j}, \dots, y_{sj})^T$ are the vectors of inputs and outputs values for *DMU* j .

Respectively let $X_j \geq 0$ & $X_j \neq 0$ and $Y_j \geq 0$ & $Y_j \neq 0$.

By using 0-1 parameters δ_1, δ_2 and δ_3 , the production possibility sets can be written into its generalized form [9],

$$T = \left\{ (x, y) \left| \begin{array}{l} \sum_{j=1}^n \lambda_j x_j \leq x, \sum_{j=1}^n \lambda_j y_j \geq y \geq 0, \\ \delta_1 \left(\sum_{j=1}^n \lambda_j + \delta_2 (-1)^{\delta_3} \lambda_{n+1} \right) = \delta_1, \lambda_j \geq 0, j=1, \dots, n+1 \end{array} \right. \right\}$$

The following four different production possibility sets are obtained by assigning different values to $(\delta_1, \delta_2, \delta_3)$:

(i) If $\delta_1 = 0$ & $\delta_2 \in \{0,1\}$ & $\delta_3 \in \{0,1\}$, then T becomes the production possibility set for the *CCR* model.

$$T_{CCR} = \{(x, y) | x \geq \sum_{j=1}^n \lambda_j x_j, \sum_{j=1}^n \lambda_j y_j \geq y \geq 0, \lambda_j \geq 0, j=1, \dots, n\}$$

(ii) If $\delta_1 = 1$ & $\delta_2 = 0$ & $\delta_3 \in \{0,1\}$ then T becomes the production possibility set for the *BCC* model.

$$T_{BCC} = \{(x, y) | x \geq \sum_{j=1}^n \lambda_j x_j, \sum_{j=1}^n \lambda_j y_j \geq y \geq 0, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, j=1, \dots, n\}$$

Theorem 1. (Bon ferroni inequality):

If the set of arbitrary events A_1, A_2, \dots, A_n constitutes a partition of the sample space S and $A_1^c, A_2^c, \dots, A_n^c$ are complements of events A_1, A_2, \dots, A_n the following rule apply:

$$P\left(\bigcap_{i=1}^n A_i^c\right) \geq 1 - \sum_{i=1}^n P(A_i) \quad (1)$$

3. Adding Competitive DMU to the Basic DEA Models

In this section, we will introduce two basic models of DEA with special condition, new DMU with stochastic inputs and outputs is added to the set of observed DMUs then information about efficiency of this DMU and changes in frontier of production possibility set of observed DMUs are interested. Suppose DMU_{n+1} with Stochastic inputs and outputs enter the previous comparative DMU system. We want to achieve results regarding the efficiency of DMU_{n+1} and impact of the amount of deterministic inputs and outputs on the old PPS frontier.

A Chance Constrained Programming model for CCR model with the assumption that DMU_o with a stochastic inputs and outputs enters the system is defined as the following:

$$\begin{aligned} \text{Min } & \theta \\ \text{s.t. } & P_r[(\theta X_o, Y_o) \in T_c] \geq 1 - \alpha \\ & \lambda_j \geq 0, \quad j = 1, 2, \dots, n \end{aligned}$$

Where X_o and Y_o are all random variables and P_r degree of probability and follow from each constraint being realized with a minimum probability of $1 - \alpha$ $0 < \alpha < 1$.

For illustrative purpose, let us assume X_o and Y_o are normally distributed with known means and variances. Thus we have the following chance constrained programming problem

$$\begin{aligned} \text{Min } & \theta \\ \text{s.t. } & P_r\left(\sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{io}, \sum_{j=1}^n \lambda_j y_{rj} \geq y_{ro}\right) \geq 1 - \alpha \quad i=1, \dots, m \quad r=1, \dots, s \\ & \lambda_j \geq 0, \quad j=1, 2, \dots, n \end{aligned}$$

Suppose:

$$A_i = \sum_{j=1}^n \lambda_j x_{ij} > \theta x_{io} \quad i = 1, \dots, m$$

$$A_{m+r} = \sum_{j=1}^n \lambda_j y_{rj} < y_{ro} \quad r = 1, \dots, s$$

Following the theorem 1:

$$P_r\left(\bigcap_{i=1}^{m+s} A_i^c\right) = P_r\left(\sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{io}, \sum_{j=1}^n \lambda_j y_{rj} \geq y_{ro}\right) \geq 1 - \alpha \quad (2)$$

From (1) and (2) we have:

$$\left. \begin{aligned} P_r\left(\bigcap_{i=1}^{m+s} A_i^c\right) & \geq 1 - \sum_{i=1}^{m+s} P(A_i) \\ P_r\left(\bigcap_{i=1}^{m+s} A_i^c\right) & \geq 1 - \alpha \end{aligned} \right\} \Rightarrow P_r(A_i) \leq \frac{\alpha}{m+s} \quad i = 1, \dots, m+s$$

We have introduced DMU_o with normally distributed, it means:

$$\begin{aligned} x_{io} & \sim N(\mu_i, \sigma_i^2) \quad , \quad y_{ro} \sim N(\mu_r, \sigma_r^2) \\ i & = 1, \dots, m \quad , \quad r = 1, \dots, s \end{aligned}$$

Then it can be written:

$$P_r(A_i) = P_r\left(\sum_{j=1}^n \lambda_j x_{ij} > \theta x_{io}\right) \leq \frac{\alpha}{m+s} \quad i = 1, \dots, m \quad (3)$$

$$P_r(A_{m+r}) = P_r\left(\sum_{j=1}^n \lambda_j y_{rj} < y_{ro}\right) \leq \frac{\alpha}{m+s} \quad r = 1, \dots, s \quad (4)$$

By (3):

$$\begin{aligned} P_r\left(\sum_{j=1}^n \lambda_j x_{ij} + \lambda_o x_{io} > \theta x_{io}\right) & \leq \frac{\alpha}{m+s} \quad i = 1, \dots, m \\ P_r\left(\sum_{j=1}^n \lambda_j x_{ij} > (\theta - \lambda_o) x_{io}\right) & \leq \frac{\alpha}{m+s} \end{aligned} \quad (5)$$

It is known that

if $X \sim N(\mu, \sigma^2)$ Then $\frac{X - \mu}{\sigma} \sim N(0,1)$

From (5) we have:

$$P_r \left(\frac{\sum_{j=1}^n \lambda_j x_{ij} - \mu_i(\theta - \lambda_o)}{\sigma_i(\theta - \lambda_o)} > \frac{x_{io}(\theta - \lambda_o) - \mu_i(\theta - \lambda_o)}{\sigma_i(\theta - \lambda_o)} \right) \leq \frac{\alpha}{m+s} \quad (6)$$

Let ϕ represented the cumulative distribution function (CDF) of the standard normal distribution and suppose $K_{\frac{\alpha}{m+s}}$ be the

standard normal value such that $\phi(K_{\frac{\alpha}{m+s}}) = 1 - \frac{\alpha}{m+s}$ then

from (6):

$$\phi \left(\frac{\sum_{j=1}^n \lambda_j x_{ij} - \mu_i(\theta - \lambda_o)}{\sigma_i(\theta - \lambda_o)} \right) \leq \frac{\alpha}{m+s} = 1 - \phi(K_{\frac{\alpha}{m+s}})$$

Finally:

$$\frac{\sum_{j=1}^n \lambda_j x_{ij} - \mu_i(\theta - \lambda_o)}{\sigma_i(\theta - \lambda_o)} \leq -K_{\frac{\alpha}{m+s}} \leq 1 - K_{\frac{\alpha}{m+s}} \Rightarrow$$

$$\sum_{j=1}^n \lambda_j x_{ij} + \lambda_o \left[\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \right] \leq \theta \left[\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \right] \quad (7)$$

Same process can be considered for (4)

Then the deterministic form of this model is:

Min θ

$$s.t \quad \sum_{j=1}^n \lambda_j x_{ij} + \lambda_o \left[\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \right] \leq \theta \left[\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \right]$$

$$\sum_{j=1}^n \lambda_j y_{rj} + \lambda_o \left[\mu_r + \sigma_r(K_{\frac{\alpha}{m+s}} - 1) \right] \geq \left[\mu_r + \sigma_r(K_{\frac{\alpha}{m+s}} - 1) \right] \quad (8)$$

$$\lambda_j \geq 0, \quad j = 1, 2, \dots, n$$

Deterministic form of BCC model for measuring the efficiency of DMU_o is:

Min θ

$$s.t \quad \sum_{j=1}^n \lambda_j x_{ij} + \lambda_o \left[\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \right] \leq \theta \left[\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \right]$$

$$\sum_{j=1}^n \lambda_j y_{rj} + \lambda_o \left[\mu_r + \sigma_r(K_{\frac{\alpha}{m+s}} - 1) \right] \geq \left[\mu_r + \sigma_r(K_{\frac{\alpha}{m+s}} - 1) \right] \quad (9)$$

$$\sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0, \quad j = 1, 2, \dots, n$$

The determination of the interval where $K_{\frac{\alpha}{m+s}}$ is changeable.

By $Y > 0, y_j \geq 0, X > 0, x_j \geq 0 \quad j = 1, \dots, n$, we have:

$$\mu_i + \sigma_i(1 - K_{\frac{\alpha}{m+s}}) \geq 0 \Rightarrow (\mu_i + \sigma_i) \geq \sigma_i K_{\frac{\alpha}{m+s}} \Rightarrow K_{\frac{\alpha}{m+s}} \leq 1 + \frac{\mu_i}{\sigma_i} \quad (10.a)$$

$$\mu_r + \sigma_r(K_{\frac{\alpha}{m+s}} - 1) \geq 0 \Rightarrow (\mu_r - \sigma_r) + \sigma_r K_{\frac{\alpha}{m+s}} \geq 0 \Rightarrow K_{\frac{\alpha}{m+s}} \geq 1 - \frac{\mu_r}{\sigma_r} \quad (10.b)$$

And by $0 < \alpha < 1 \Rightarrow 0 < 1 - \alpha < 1$:

$$1 - \alpha > 0 \Rightarrow \alpha < 1 \Rightarrow \frac{\alpha}{m+s} < \frac{1}{m+s} \Rightarrow 1 - \frac{\alpha}{m+s} > 1 - \frac{1}{m+s}$$

$$\Rightarrow \phi(K_{\frac{\alpha}{m+s}}) > 1 - \frac{1}{m+s} \Rightarrow K_{\frac{\alpha}{m+s}} > \phi^{-1}(1 - \frac{1}{m+s}) \quad (10.c)$$

By (10.a), (10.b), (10.c), $K_{\frac{\alpha}{m+s}}$ is obtained.

4. Numerical example

In this section, we work out a numerical example to illustrate the efficiency of competing DMU with stochastic inputs and outputs that enter our system and suppose the inputs and outputs are normally distributed with known means and variances(given by Decision-maker) deterministic model and its efficiency are interested.

Example: Consider the four DMUs with single inputs and single outputs as defined in Table 1:

Table 1
Data set for the numerical example

DMU	A	B	C	D	E
Input	2	5	2	7	$x_E \sim N(2,4)$
Output	4	8	2	8	$y_E \sim N(8,9)$

By adding DMU_E with stochastic inputs and outputs and using (2) for CCR model, we have:

$$\begin{aligned} \text{Min } & \theta \\ \text{s.t. } & P_r \left[\begin{array}{l} -2\lambda_1 - 5\lambda_2 - 2\lambda_3 - 7\lambda_4 - x_{15}\lambda_5 + x_{15}\theta \geq 0 \\ 4\lambda_1 + 8\lambda_2 + 2\lambda_3 + 8\lambda_4 + y_{15}\lambda_5 \geq y_{15} \end{array} \right] \geq 1 - \alpha \quad (11) \\ & \lambda_1, \lambda_2, \dots, \lambda_5 \geq 0 \end{aligned}$$

The deterministic form of (11) is:

$$\begin{aligned} \text{Min } & \theta \\ \text{s.t. } & 2\lambda_1 + 5\lambda_2 + 2\lambda_3 + 7\lambda_4 + \lambda_5[2 + 2(1 - K_{\frac{\alpha}{2}})] \leq \theta[2 + 2(1 - K_{\frac{\alpha}{2}})] \\ & 4\lambda_1 + 8\lambda_2 + 2\lambda_3 + 8\lambda_4 + \lambda_5[8 + 3(K_{\frac{\alpha}{2}} - 1)] \geq [8 + 3(K_{\frac{\alpha}{2}} - 1)] \\ & \lambda_1, \lambda_2, \dots, \lambda_5 \geq 0 \end{aligned}$$

The deterministic form of BCC model for measuring the efficiency of DMU_E is:

$$\begin{aligned} \text{Min } & \theta \\ \text{s.t. } & 2\lambda_1 + 5\lambda_2 + 2\lambda_3 + 7\lambda_4 + \lambda_5[2 + 2(1 - K_{\frac{\alpha}{2}})] \leq \theta[2 + 2(1 - K_{\frac{\alpha}{2}})] \\ & 4\lambda_1 + 8\lambda_2 + 2\lambda_3 + 8\lambda_4 + \lambda_5[8 + 3(K_{\frac{\alpha}{2}} - 1)] \geq [8 + 3(K_{\frac{\alpha}{2}} - 1)] \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ & \lambda_1, \lambda_2, \dots, \lambda_5 \geq 0 \end{aligned}$$

So from the intersection of (10.a), (10.b), (10.c), $0 < K_{\frac{\alpha}{2}} \leq 2$. By Supposing $\text{Range}(K_{\frac{\alpha}{2}}) = 0.1$

Now, we apply our method to the data set. The results are provided in Table 2. As we can see in Table 2, for DMU_E The Deterministic Inputs and Outputs of DMU_E and Objective Function that is measured with CCR and BCC Model. The last column shows that in which condition (value of α) DMU_E is MPSS.

Fig. 1. Old and New Efficient Frontiers for the Example
The old and new efficient frontiers to this data set are depicted Fig. 1. As the figure shows, DMU_E after 4 stage with special value of α will be Most Product Scale Size (MPSS).

5. Conclusion

Randomness in problem data poses a serious challenge for solving many linear programming problems especially in

DEA. The goal is to compute the least amount of α in which the discussed DMU is efficient for the first time.

Table 2

The Deterministic Inputs and Outputs of DMU_E and Objective Function that is measured with CCR and BCC Model.

$K_{\frac{\alpha}{2}}$	$1 - \alpha$	I_1	O_1	θ^*_{BCC}	θ^*_{CCR}	MPSS or Non MPSS
0.1	0.0796	3.8	5.3	0.783	0.697	Non MPSS
0.2	0.1586	3.6	5.6	0.889	0.778	Non MPSS
0.3	0.2358	3.4	5.9	1	0.868	Non MPSS
0.4	0.3108	3.2	6.2	1	0.969	Non MPSS
0.5	0.3830	3.0	6.5	1	1	MPSS
0.6	0.4514	2.8	6.8	1	1	MPSS
0.7	0.5160	2.6	7.1	1	1	MPSS
0.8	0.5762	2.4	7.4	1	1	MPSS
0.9	0.6318	2.2	7.7	1	1	MPSS
1.0	0.6826	2.0	8.0	1	1	MPSS
1.1	0.7286	1.8	8.3	1	1	MPSS
1.2	0.7698	1.6	8.6	1	1	MPSS
1.3	0.8064	1.4	8.9	1	1	MPSS
1.4	0.8384	1.2	9.2	1	1	MPSS
1.5	0.8664	1.0	9.5	1	1	MPSS
1.6	0.8904	0.8	9.8	1	1	MPSS
1.7	0.9108	0.6	10.1	1	1	MPSS
1.8	0.9282	0.4	10.4	1	1	MPSS
1.9	0.9426	0.2	10.7	1	1	MPSS

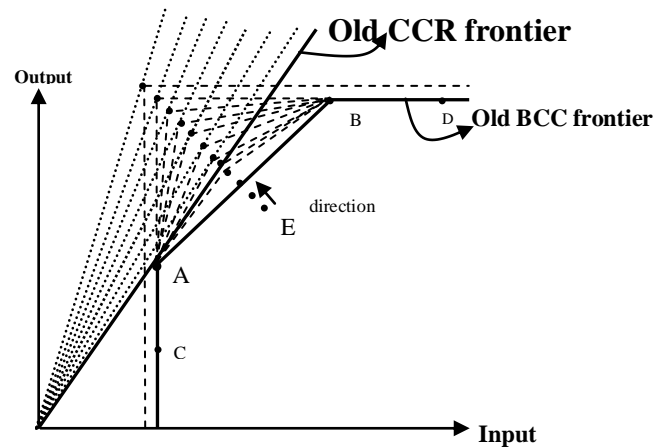


Fig. 1. Old Efficient and New Efficient Frontiers for the Example

Two cases might happen either the DMU locates on the frontier that the other DMU's had made before the use of comparative DMU. In this case the efficiency of other DMU's is preserved. There may be just a ranking order of the DMU's (in the CCR model). or as in the second case this DMU might not be located on the old frontier. In this case the PPS frontier changes and gains new shape. We can 'not say any thing about the DMU's that were efficient before adding the competitive DMU and their efficiency must be measured again.

Probability of either two cases is dependent upon the amount of mean and variance of the DMU's available for us by the DM.

6.Acknowledgment

Authors sincerely thank **Dr. F. Hosseinzadeh Lotfi**, for his most valuable comment and encouragement, which stimulated a significant revision on the description of research motivation, explanation and conclusion of the paper.

References

- [1] Lai, Y.J, C.L, Hwang, "Fuzzy mathematical programming, *Applied Mathematics and Computation*" (1991) 268-275.
- [2] Dantzig, G. B, "Linear programming under uncertainty. *Management*" (1955) *Sci.* 1: 197–206.
- [3] Beale, E. M. L, "On minimizing a convex function subject to linear inequalities." *J. R. Stat. Soc.* (1955) *B17*: 173–184.
- [4] Beale, E. M. L, "The use of quadratic programming in stochastic linear programming.", (1961) Rand Report P-2404, The RAND Corporation.
- [5] Tintner, G, Stochastic linear programming with applications to agricultural economics. In Antosiewicz H. (ed) *Proc. 2nd Symp. Linear Programming*, volume 2, (1955) pages 197–228. National Bureau of Standards, Washington D.C.
- [6] Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision-making units. *European Journal of Operations Research* 2, 429–444.
- [7] Charnes A, Cooper WW. Deterministic equivalents for optimizing and satisfying under chance constraints. *Mgmt Sci* 1963;11:18-39.
- [8] Avriel, M. and Williams, A, The value of information and stochastic programming. *Oper. Res.* (1970) 18: 947–954.
- [9] Wei, Q.L, H., Yan, Characteristics and structures of weak efficient surfaces of production possibility sets, *European Journal of Operations Research* 2, (2007) 3-4 .

Concentration on Business Values for SOA-Services: A Strategy for Service's Business Values and Scope

Bashir Ahmad¹, Sheikh Muhammad Saqib¹, Muhammad Zubair Asghar¹, Muhammad Ahmad Jan¹ and Shakeel Ahmad¹
¹Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

Abstract—Markets simply buy that what industry produce. Service Oriented Architecture (SOA) enrolment in market works on this formula. Produced services in SOA work for many organization. Is it possible to incorporate business values of different organization in same service? Different authors suggest different business value's practices. But still confusion is in same place that "people fail to explain SOA business value". Although SOA business values is some time very hard to find because there will no a particular user of service. Here author evaluated the different business value's practices and suggest a strategy for extracting SOA business values. Proposed strategy can help users to get themselves familiar with upcoming new services. Hence with proposed strategy, architect can design the service according to user needs.

Keywords- SOA; Business Values; Service;

I. INTRODUCTION AND PROBLEM DEFINITION

For designing any service, designer must have knowledge about what target they want to achieve? Without business values, how this target can be achieved? Since in SOA mostly are enterprise level applications, then how these values can be determined and from where? Suppose we take the services for an antivirus.

If architects design an antivirus for different users then without knowing following some questions, how they design different services for antivirus. How computer was affected? On what time computer was affected? Affected computer was connected with internet, intranet or network? Computer was affected by the execution of any file? How you executed such type of file voluntarily or involuntarily? If yes, then which was the type of file? There are some types of questions for antivirus services. Designer can explore these questions, but their answers can be carried out from business values. From where these business values are determined because there are different organizations which can use these services. And also architect must design these services by keeping the fact in mind that these should be loosely coupled. In Service Oriented Architecture, Services are served on end point in the network and communicates different messages according to its specification. This specification is described its constraints. The user of the service has little or no knowledge about how the service is implemented or how it is provided [3]. It is almost right, then how architect can design a service for an organization without capturing business value. Because business values contain all those parameters through which health and wealth of a firm can

be determined. But SOA based projects are not for just single organization, so architects can make the global view of business value for a service.

Since SOA application is of enterprised, so different users from different organization can use this service. Two questions arise here:

- A service for specific purpose for all organization can have same business values?
- Can a service contain the different business objectives from all organization to specific purpose?

Different author suggest different approaches for business value in SOA. Now we determine whether their approaches can lead to solve these two questions.

A bridge is required between business and technical world for enterprise applications. Business needs can easily carry out if IT infra structure should follow the business structure. Through SOA, this goal is possible if organization adopts methodology and processes specific to SOA structured design and analysis. Analysis means a process that often begins with information gathering. Here author suggest that SOA will be started within an organization with an objective of business. Through analysis, these Services are classified into Service Layers: Service Entity (Product, Customer, Service, Invoice ...) Service Task (Order Management, Business processing) Service Utility (Non Business Centric,) [2]. If we consider service entity, service task, service utility as business values and these are carried out from an organization then question is arised that this service with the said business values can be plugged-in with other applications for other organization?

The discussion sessions about the business values of SOA has been made between audience and the group of architects. These sessions have been very informative for those groups of architecture. But there was not any good answer that can satisfy everyone [4]. Many points are concluded from those discussions but major point is. "The underlying IT infrastructure which supports those business processes needs to be flexible and capable of adapting to change" [4].

Concluded points are very beneficial for SOA business values, but still problem is arised how these points can be achieved?

Business Process Management BPM can solve some issues about SOA business values [6]. BPM is a logical approach which can improve business processes of company. For greater business agility, SOA enables

different services for recombining business processes. If SOA is considered as road then BPM will be considered as a car that can take something for a business from registry. "SOA by itself can be a hard sell since it can be difficult to explain the value in concrete, understandable terms, and thus makes it difficult to convey its value". BPM can solve this limitation because it can convey SOA values to specific business process [5]. Since BPM can determine specific service according to user needs. So SOA buying process problem can be solved with BPM, but for architecting new service how business values can be carried out from different dimensions.

II. SERVICE ORIENTED ARCHETECTURE

Basically SOA is a combination of collected services, where each service can communicate with other services. This communication can be simple data passing or coordinating some activities with other services. In SOA, service is an independent unit for deployment.

SOA enhances reuse and linked with computational resources. SOA has flexibility and makes cost effective changes according to current situation of market. It recommends the use of existing assets and interconnection between the assets. SOA contains different characteristics:

- By adopting XML-Schema, these services can communicate with each other.
- Applications can take a service from pool of services' repository with help of SOA supervisor.
- Each service contains quality with respect to security recruitments, policies, authentication, authorization and reliability.

As services in SOA are loosely coupled, then it requires some techniques for forwarding the services to consumers. Share out some simple and small interfaces between the participating software agents. These interfaces should entertain to all providers and consumers globally. Loosely coupled features can be easily adopted by using the above technique. Many people focus on development of flexible system with the combination of loosely coupled services and systems should be higher in quality and cheaply. This can be done by getting the business values from business [7] [8].

III. STRATEGY TO DETERMINE BUSINESS VALUES FOR SOA-SERVICES

With the greater collaboration between business users and IT, development efforts can be more effectively. The best approach for business values is center of excellence (COE). "The COE should have representatives from the business side and the IT side, but the business needs should be addressed first. What problem really needs solving? Then figure out how." [1]. Business values can easily point out if business addressed first. But when architect design a new service for many organization, then from where they can extract these values? Since these are global services for all organizations (just like antivirus) then a survey must be conducted from some organizations to carry out some business values and their scope. Whole procedure is shown in figure below.

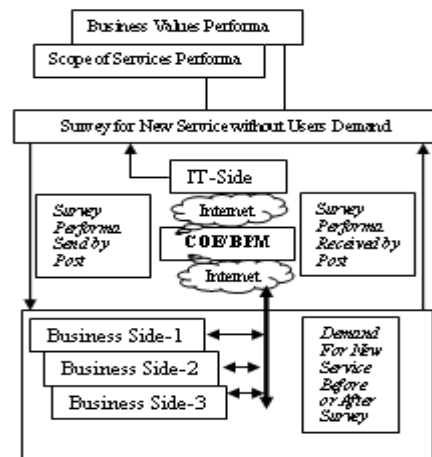


FIG-1: EXTRACTING SOA BUSINESS VALUES.

Above figure is showing that business side users can send a request for their own service through COE or BPM. But when IT-side wants to design new services after exploring some issues, they should conduct a survey from different organization to get out its business values. After conducting this survey, market can have how know about new services. So before coming, new service can take a place in market container. Survey will determine business values and scope of different services. Following table will be used to determine business values.

TABLE-1: BUSINESS VALUES FOR SERVICES

Pre filled by Designer		Filled by Organization		SOA Business Values	
				Pre filled by Designer	Filled by Organization
Service NOs	Service Name	Service Category	General	Designer Observations Requirements	Client's Requirements
			Specific		
			More		
			Specific		
Service-1					
Service-2					
Service-3					

In above Table-1, IT side will filled the services name and designer observation requirements. Business side will filled the different requirements according to their needs and will confirm the service category. Service category contains only one value from general, specific or more specific. Meaning of these options is: General: Service can be used for all users of an organization. Specific: Service can be used only particular category of users. More Specific: Service can be used for Specific only one user.

After getting the above information from different business sides, IT side will check three conditions to determining scope. Table-2 will be used for determining the scope of different services. Following three points can be helpful for exploring the scope. 1. Use in Only single Organization: Means Services can not be used besides single organization. 2. Use in Same type of organization: Means Service can be used in same type of organizations. 3. Use in other type of organization: Means Service can be used in all type of organization.

TABLE-2: SCOPE OF SERVICES

Service No	Scope of Services					
	Use in Only single Organization?	Y	Use in Same Type of Organization?	Y	Use in other Type of Organization?	Y
		N		N		N
Service-1	YES or No (No means service can be used here and other)		If Yes Then Total Organization??		If Yes Then Total Organization??	
Service-2						
Service-3						

After getting the scope of service for using in different type of organization and business values about services, architect can easily design the structure of a service in loosely coupled manner, which will be easily plugged-in in any business IT.

IV. CONCLUSION:

An architectural and integration framework SOA, which will work on loosely coupled services and integrated them in new or legacy system [5]. Since SOA is modeling the business process and without business values, how a process is modeled. In SOA exploring business values is passing to confusions, here author suggest a strategy which support to extract SOA business values and an unknown service can make a place in heart of market.

V. REFERENCES:

- [1] Judith Lamont, Leveraging SOA for Business Value, 2007
<http://www.kmworld.com/Articles/Editorial/Feature/Leveraging-SOA-for-Business-Value--40506.aspx>
- [2] Sadi Melbouci, Methodology to deliver Service Oriented Architecture, SOA, 2010
<http://community.jboss.org/wiki/MethodologytodeliverServiceOrientedArchitectureSOA>
- [3] Andy Tan, Why we need to pay attention to the SOA: COMPUTERWORLD Year-End Bumper Issue 2005
URL: <http://www.fwsi.org/rt-06002.pdf>
- [4] David A. Chappell , Extracting Business Value from SOA, 2005
http://www.oreillynet.com/xml/blog/2005/02/extracting_business_value_from.html
- [5] Mark Cooper, "Business Process Management (BPM) Definition and Solutions", 2007
http://www.cio.com/article/106609/Business_Process_Management_BPM_Definition_and_Solutions?page=4#soa
- [6] Judith Lamont, Leveraging SOA for Business Value, 2007
<http://www.kmworld.com/Articles/Editorial/Feature/Leveraging-SOA-for-Business-Value--40506.aspx>
- [7] Service Oriented Architecture, August 2007
<http://www.javajazzup.com/issue2/page28.shtml>
- [8] Neil Ward-Dutton; SOA's Business Value; 2006
<http://www.ebizq.net/topics/soa/features/7338.html?page=2>

LOAD-BALANCING GEOGRAPHIC ROUTING ALGORITHM (ELBGR) FOR WIRELESS SENSOR NETWORKS

Nazia Perwaiz

Department of Computer Engineering, NUST College of
Electrical & Mechanical Engineering, National
University of Sciences & Technology, Islamabad,
Pakistan.
missnazia@ceme.nust.edu.pk

Dr. Muhammad Younus Javed

Department of Computer Engineering, NUST College of
Electrical & Mechanical Engineering, National
University of Sciences & Technology, Islamabad,
Pakistan.
myjaved@ceme.nust.edu.pk

Abstract: *The major challenges faced by wireless sensor networks are energy-efficiency and self-organization. A thorough literature study of routing in WSNs shows that there exist so many routing protocols for WSNs, each of which has the common objective of trying to get better throughput and to extend the lifetime of the sensor network. In this research work the Location-based or Geographic routing in WSN is mainly focused for energy issues and a location based WSN routing protocol ELBGR (Energy aware & Load Balancing Geographic Routing) is proposed that features the energy efficiency and self organization of the wireless sensor networks. This protocol extends the lifetime of the network and balances the energy consumption of the nodes within the network. ELBGR considers neighbor's energy levels, packet reception rate and the locations of the nodes for data forwarding purpose. Each node knows geographic location, energy levels and PRR of its neighbors. The proposed algorithm selects a set of relative optimal nodes from all neighbors called Forwarding Nodes Set (FNS) in the first phase and in the next phase from FNS, the Optimal Forwarding Node (OFN) is finally selected for forwarding purpose. The proposed algorithm balances the energy levels among all the neighbors. A comparison has been made between pre-existing routing algorithms Greedy routing, EAGR, EEAR, HHEAA and the proposed one ELBGR. The Simulation results (in OMNET++) show that the proposed algorithm gives better performance in terms of higher success rate, throughput and less number of dead nodes and it effectively increases the lifetime of the sensor networks.*

Keywords: WSN; Geographic routing; Energy efficient; Load balancing

I. INTRODUCTION

The technological progress in the embedded systems emerged a new network class called Wireless Sensor Networks (WSN) [1]. A WSN consists of many small autonomous systems, called sensor nodes or motes. Sensors are the devices that make sense to some physical change for which they are deployed for different

applications, communicate wirelessly and cooperate with the neighbor nodes to route the sensed information towards the destination. This type of communication has different characteristics as compared to cellular networks or single hop wireless networks as they don't rely on a fixed infrastructure.

Many routing strategies have been introduced and developed for the WSNs. These strategies used in WSN are different from wired networks routing mechanisms. The schemes used for routing in WSN must not ignore the unique inherent features of the WSNs, the main critical issue in WSNs routing strategies development is to deal with the energy constraints and to cope with the nodes status changes (e.g. failure) that occur suddenly resulting unpredictable changes in the network topology. The communication protocols governing the network must be able to cope with all topological changes without human intervention. Most of the nodes are too far away from the sink node (node that finally collects the sensed information) to communicate directly. Intermediate nodes are hence used to relay the message, that's why this type of communication is also called ad hoc multi-hop communication [19].

The major purpose of this research is to develop an energy aware and energy efficient geographic routing algorithm for the WSNs that can play an important role in maintaining the energy balance within the network causing the prolonged life-time of the network. A geographic routing mechanism, ELBGR (Energy & Load Balancing Geographic Routing) is proposed that is much efficient for energy efficiency, it splits the work load equally among all network nodes and avoids the holes formation than the greedy forwarding in WSNs. The proposed algorithm depicts the node that will act as the relaying node among all one-hop neighbors of the sender node on the basis of some relative measures rather than some specific threshold values. This way the over burden of relaying on some nodes is abandoned and the load is divided among all the network nodes, resulting energy efficient long life network.

The paper is organized as follows. The Related work is presented in section 2. The research motivation, objectives, assumptions and the proposed algorithm ELBGR is described in Section 3. In Section 4 the

proposed algorithm's design and implementation along with the algorithm and flow chart is presented. The simulation results, comparison of proposed algorithm with pre-existing algorithms Greedy, EEAR, EAGR, HHEAA and the results analysis is explained in Section 5. Conclusions along with some suggestions for future research are provided in Section 6.

II. RELATED WORK

In the location-based or geographic routing [7], the location information is basic requirement for the neighbor nodes identification and routing. Many routing algorithms are defined that use only local information of neighbors for route determination, these are much energy efficient than the algorithms using global routing table, that depict the whole picture of the network.

Greedy Perimeter Stateless Routing (GPSR) [8] is designed to minimize the number of hops, which follows greedy forwarding algorithm along with the perimeter forwarding strategy to achieve successful routing. Using greedy forwarding when optimal next-hop nodes are not available, the perimeter forwarding algorithm is used on the planar graph to choose the next-hop. Geographic adaptive fidelity (GAF) [10] is a geographic and energy-aware routing protocol, used both for ad-hoc networks and WSNs. It decreases the redundancy in the network and turns off all unused nodes of the network. Greedy other adaptive face routing (GOAFR) [11] is a geographic routing scheme that uses the greedy routing along with the face routing called other adaptive face routing. When GOAFR reaches the local minimum point it adopts the face routing (FR) mode to get routing efficiency for both worst case & average case scenarios. GOAFR protocol is much better in energy efficiency. It's time delay in data delivery and network lifetime is also much better. Geographic and energy aware routing (GEAR) protocol [12] uses the nodes' location information and the remaining energy level to select the neighbor node with the least overall overhead. The estimated cost is used for simple routing and based on nodes' remaining energy levels and its distance to the destination node & the learned cost is used for routing around the holes. GEAR uses the recursive geographic forwarding (for dense network) and the restricted blind flooding (network is not dense) to disseminate the packet within the region. GEAR is not highly

energy efficient, has low latency and QoS. It performs better for immobile networks but inefficient for mobile networks. Energy aware greedy routing (EAGR) [15] uses the location information of nodes and their powers available for routing purpose. In EAGR, high energy nodes are used for forwarding and packets are dropped when no neighbor is alive to forward the data. EAGR protocol is much more energy efficient and balances the network loads very well. Holes Healing Energy Aware Algorithm (HEEAA) [16] is location based and energy aware routing protocol that works on average energy and distance of nodes to overcome weak node problem in WSN. HHEAA has high throughput with reliable packet delivery and long lived network. Efficient Energy Aware Routing (EEAR) [17] algorithm is a location based & a power-aware routing technique, it uses efficient energy aware routing mechanism to choose the neighbor node that has sufficient power-level and meets the distance criteria to determine the receiving node for forwarding the packet. EEAR gives higher packet delivery rate, less energy consumption with the maximum network life time as compared to traditional routing mechanisms for the wireless sensor networks.

III. ENERGY AWARE & LOAD BALANCING GEOGRAPHIC ROUTING (ELBGR)

A. MOTIVATION

Many of the geographic routing techniques designed & developed for the WSN uses the basic idea of greedy technique for forwarding the data, the minimum distant node (from destination) is used for this purpose to choose shortest path and early delivery of the packet. Hence the main problem in routing algorithms designed and used for the WSNs is that how much energy efficient [25] they are and their role to enhance the life of the network. Very important feature of the WSN is that the nodes have small sized batteries, limitedly-powered & these batteries cannot be changed in practical therefore it is required to use some ways to save the batteries power to prolong the network life so the WSN is taken as energy-constrained. Because batteries cannot be replaced, communication protocols ought to be as energy-efficient as possible. Energy-efficiency is of utmost importance, as it has very visible impact on the network lifetime. It is required to develop some energy efficient routing mechanism which maintains a balance to the overall network power and holes appearance be prevented consequently.

The motivation behind this research work is to design an energy-aware & energy-balancing geographic algorithm

for WSNs which will be simple, easy to implement and efficient in terms of energy consumption.

B. RESEARCH OBJECTIVE

Greedy mechanism used in many geographic routing has a major drawback that the construction of the routes is based only on the node's distance from the destination and the same group of nodes is repetitively used for sending data from the source to the destination if source & destination are constant. The energy-level of a node is not taken into account for data transmission and even if the node has very low energy, the packet can be forwarded to it resulting in the breakdown of that optimal path. It puts bad effect on network connectivity as there may exist some of the nodes in networks which depend on only these dead nodes for routing the packet. In this research many geographic/location based routing strategies for WSNs are studied for complete understanding. To address the above mentioned problems an energy-aware and load balancing greedy routing scheme ELBGR (Energy & load balancing Geographic Routing) is designed, implemented and evaluated through simulation. The comprehensive simulation results are compared with pre-existing relevant routing protocols for its energy efficiency, successful data transmission and for some other parameters. ELBGR is a simple, easy to implement and energy-efficient algorithm, it splits the work load equally among all network nodes and avoids the holes formation resulting energy efficient long life network.

C. ASSUMPTIONS FOR ELBGR

For the designing and implementation of the ELBGR routing algorithm some assumptions are taken. Sensor nodes are considered to be static or immobile having fixed coordinates. The location of the nodes is determined by some kind of GPS system (some central-location database is used for this purpose). The energy levels and the PRR values of the nodes are known by each node; initially both values are set 01 for each node. The topology used is irregular random topology as in real the sensors are deployed in random style. Single destination node is considered with already known location by each node. The limited-size buffers/ queues are used at each node for containing incoming and outgoing message packets. Packets size is taken fixed for the proposed system implementation.

D. DESCRIPTION OF ELBGR

The major role of the proposed algorithm is the avoidance of holes formation in the network by equally distributing the load among the nodes, no single node depletes its energy very soon as relatively optimal node is selected for forwarding purpose. Each node know about its own distance from sink node and its all neighbors' distance from sink node, energy level and PRR value ($PRR = \text{total packets sent by node A} / \text{total}$

packets received by node A). Based on this all information each node contains, the ELBGR extracts the relatively optimal node among all neighbors, first of all finds the average distance of all nodes' distances from the sink/destination node, & then it calculates the average energy and average PRR of all neighbor nodes. After taking these measures, this algorithm selects the neighbor nodes for FNS which have average or greater remaining energy as well as average or greater PRR value. It is obvious that greater energy nodes selection for forwarding purpose prevents the high rate of nodes death in the network. Nodes PRR values are important in respect of the holes avoidance, if a node forwards all the packets its PRR value doesn't change (remains 1, initially set value) and the routing decision is almost energy dependent. But when a node drops the packet due to any reason i.e. lower energy level or queue fill, the packet reception rate lowers down and for next packet transmission that node is avoided due to low PRR value (lower than average PRR value of all neighbors), resulting the avoidance of the hole. Neighbors having energy and PRR below average value are ignored so that the balance of energy can be maintained among all neighbors and this way of selection prevent some nodes to be selected frequently resulting in energy depletion. Therefore FNS contains only nodes having greater energy and PRR values. In next phase ELBGR considers only FNS for decision making and selects the node nearest to the target/ destination node from the set of nodes that lie in FNS. After choosing the optimal forwarding node (OFN), the packet is sent to it, this current node then makes further decision for next forwarding node among its neighbors using ELBGR mechanism. This process continues until the destination node is reached. When packet is sent to a neighbor, the energy used suppose 0.001joule is deducted from its current energy level, this way the remaining energy value and current PRR value of each node is updated after each time when it forwards the packet. The new values are stored in the node's memory. For next transmission if any neighbor node considers it for forwarding, it will check its energy and PRR values, which are if less than average values then that node will not be considered for selection in FNS and the nodes which are not selected previously or having greater energy and PRR values are selected in FNS. Using ELBGR routing scheme the network work load is taken as the data packets which are needed to be sent to the destination or target node, is equally distributed (all nodes are considered equally for forwarding purpose) among all nodes and almost an energy balance is also maintained among nodes. So the path selected may not be the similar every time & the traffic is spread over many nodes rather than to the specific nodes only. As a result the sensor network is utilized for maximum time with greater throughput in energy efficient and load balancing way. ELBGR algorithm tries its best to

prevent the formation of holes, but as it is unavoidable due to energy constraints but this algorithm avoids the routing of packet towards holes considering PRR values, using ELBGR a very little number of holes can be seen in the network even after a long time span of network

utilization when almost all nodes of the network have consumed almost their whole resources.

IV. DESIGN & IMPLEMENTATION OF ELBGR

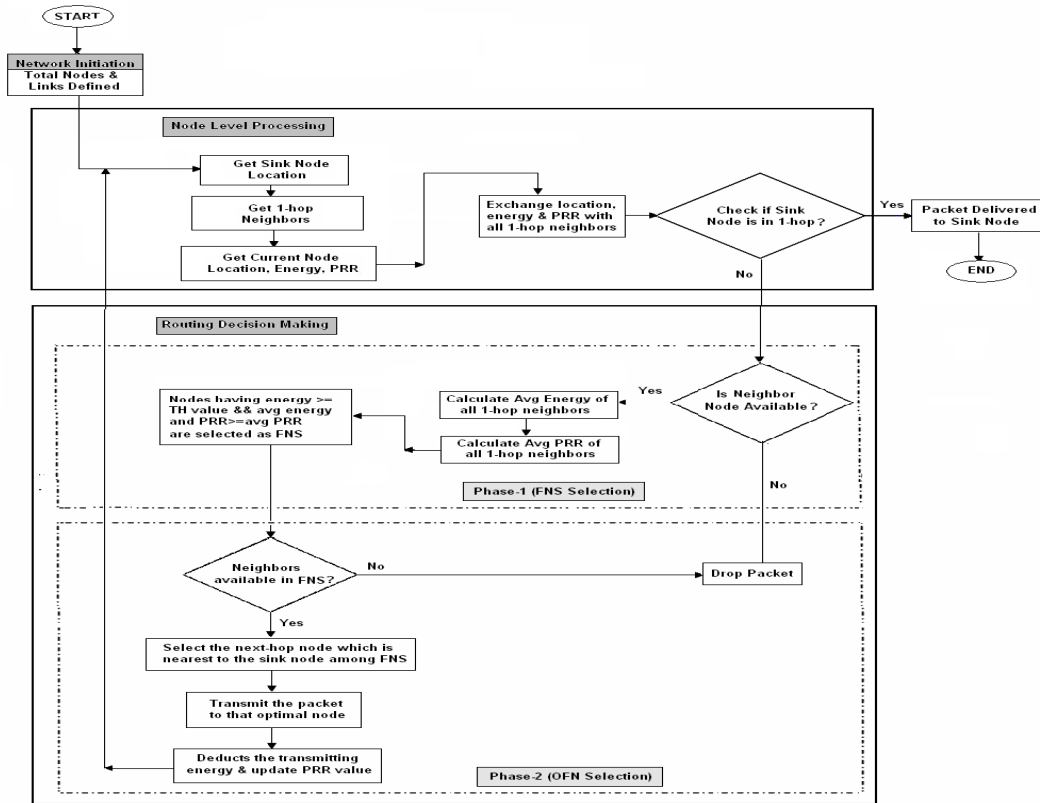


Figure 1. Flow chart of ELBGR

5. SIMULATIONS AND ANALYSIS

A. THE PROPOSED SYSTEM MODULES

The proposed system is implemented on OMNET++ Simulator. OMNET++ works on the modules system;

four different modules defined in our proposed system are network generator, Route generator, proposed ELBGR algorithm and the router module.

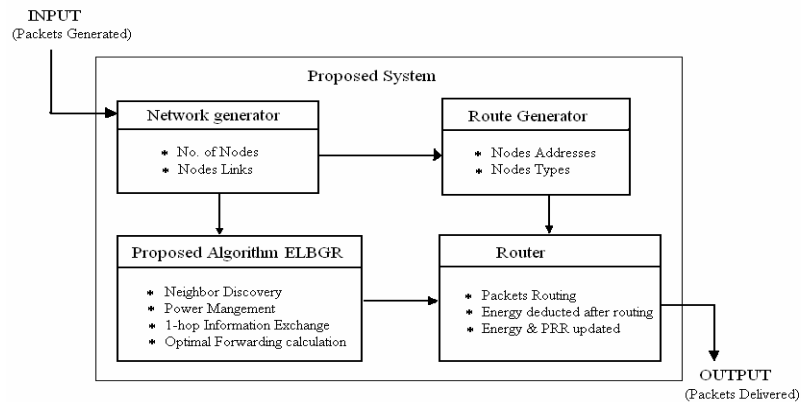


Figure 2. Proposed System Block Diagram

The *Network generator* module generates the network, the total number of nodes is defined in this module, the packet sending rate and time delay between packets transmission are also defined as parameters of this module. The *Route Generator* uses the addresses of the nodes for determination of the sending/ source node & sink/ destination node for data packet sending and receiving respectively. The *Proposed Algorithm ELBGR* is very important module that chooses the next forwarding node. The proposed algorithm assumes that all nodes know their own geographic-locations, their energy-levels and current PRR values & it implements the proposed algorithm. The *Router* module is also very important in its functionality. It performs actual routing of packet. When the packet is sent towards the neighbor node, this module subtracts the energy from the node's current energy and re-evaluates the PRR value and updates all neighbors about it. Finally it shows output of simulations as the successfully delivered-packets, dropped-packet, node's current power-level, node's current PRR value and the status of the node.

B. ELBGR SIMULATION MODEL

ELBGR Simulation model is designed in OMNET++. Different numbers of nodes exist in the network so that the simulation results may be evaluated for variable-sized network having variable number of nodes. All evaluation metrics are checked for different models with different number of nodes (different network size). In this simulation, the nodes' locations of network are taken randomly without any predefined criteria and irregular random topology is used as in real conditions the WSN sensor nodes are also deployed randomly. As per assumptions already described it is clear that the nodes are static and don't ever change their position. The position of each node is determined by the node itself using a central location database, this central database also inform each node with sink/ destination node location. In ELBGR system, the sink node is fixed and pre-defined. One node is declared as target node, as each node have knowledge about its location, each node itself measures it's distance from the target/ sink node (as for number of hops). In start each node has allocated same energy level i.e. 01 Joule and same PRR value i.e. 1. Before the simulation starts, each node has information about Node's location, Sink node location, Node's distance from sink node, Node's energy level & Node's PRR value. This all information is to be exchanged among all 1-hop neighbors for maintaining a local-table of neighbor's info, which will be used for routing decision making later. In the start a threshold (TH) energy-level is defined & the nodes whose energy-levels are lower than threshold value are considered as dead nodes. Each node has a limited size buffer, a fixed size queue is defined which is used to store incoming / outgoing packets temporarily. In this

simulation model, the packets used are of fixed size 562 bytes, different number of packets are generated by some source node at some pre-defined time delay, this time delay can be changed to analyze its effect on the proposed algorithm's performance. Total simulation time is also pre-defined. The simulation process is repeated with different number of packets generated and with different numbers of nodes to evaluate the proposed algorithm. A sample network designed in OMNET++ with 90 numbers of nodes is shown below

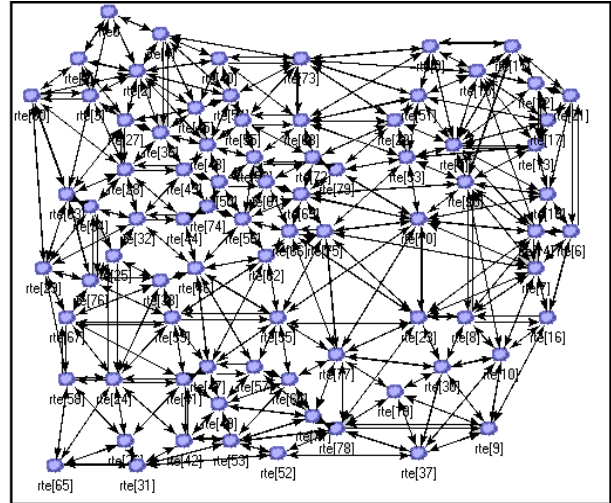


Figure 3. Sample Network with 90 Nodes

C. NETWORK INPUT PARAMETERS

For this simulation various input parameters are defined which are considered same for both routing strategies (Greedy Routing and ELBGR routing) that are compared for evaluation of the proposed one's performance.

Parameters	Value
Network Size (Number of nodes)	Variable for each scenario
Traffic type	Constant (1pkt/microsecond)
Nodes type	Static/ Fixed
Topology	Irregular random topology
Data packet size	562 byte
Buffer/ Queue size	10
Initial Energy level of each node	1 Joule
Energy Threshold value	0.1 Joule
Transmission energy	0.001 Joule
Initial PRR value	1
Number of receiver	1 (Predefined)

Table 1. Input Parameters

The network size is variable for each scenario as this research focuses that the proposed algorithm ELBGR scales very well and performs better with different sized network having different number of nodes. Yet for each scenario both the Greedy and ELBGR routing algorithms are evaluated for comparison purpose.

D. EVALUATION METRICS

To make evaluation for the proposed routing scheme's performance, some evaluation metrics are analyzed for some pre-existing algorithm greedy routing, EAGR, EEAR, HHEAA and the proposed ELBGR routing mechanism i.e. *Packets delivered, Packets Dropped, Success rate, Alive Nodes, Dead Nodes and Throughput*.

E. SIMULATION SEQUENCE OF ELBGR

As defined in input parameters before simulation starts each node has total energy equal to 1Joule. Initially each node creates a local-table of its neighbors containing all 1-hop neighbors along with their current energy levels (initially 1J), PRR values and their distance to the destination. When a node is used for transmission purpose and its energy level or PRR value changes then it informs its 1-hop neighbors to update their local tables. Nodes initiate the sending of the data to the destination by selecting a neighbor nodes set (FNS, Forwarding Nodes Set) on the basis of ELBGR algorithm. Using ELBGR, the current node first of all calculates the average remaining energy and average PRR of its all neighbors. Initially when all nodes have same energy levels and PRR values (equal to the average energy level and average PRR value respectively), ELBGR works like greedy scheme, the data packets are sent to the node closest to the sink node (minimum hops to the target node is considered the nearest or closest node), the process carries on until the arrival of the packet to the destination. After utilizing a node for forwarding purpose, its energy & PRR values are updated and exchanged with neighbors. For the next packet from the same source which is to be forwarded towards the same destination, the sending node will have same neighbor nodes but it will not use the previously selected neighbor for forwarding purpose as the nodes that have been used in previous packet forwarding have updated their local values, energy and PRR values are now less than the average energy and PRR, due to this those nodes are not selected in next transmission, some other node will be selected now. This is the beauty of ELBGR that it distributes the load among all neighbors by screening out previously used nodes; this procedure has great impact on the overall energy consumption of WSN and on the network lifetime. Secondly ELBGR doesn't use any specific values, rather it uses relatively optimal node among all neighbors, this way it utilizes almost all resources of the

nodes hence the network remain alive until almost complete consumption of all network resources.

Along with the greedy algorithm, we have compared the evaluation parameters of the ELBGR with some other proposed algorithms that proved themselves better in performance than simple greedy routing i.e. EAGR, HHEAA and EEAR.

A comparison among above mentioned algorithms is shown below for constant values of number of nodes, number of packets generated and time duration i.e. 90 nodes, 45000 number of packets generated and 500 seconds.

Parameters Evaluated	ELBGR	GREEDY	EEAR	EAGR	HHEAA
Packets Delivered	44730	26008	43499	42666	36615
Packets Dropped	270	18992	1501	2334	8385
Throughput (pkts/sec)	89.46	52.02	86.99	85.33	73.23
Alive Nodes	89	85	89	89	86
Percentage Alive (%)	98.8	94.4	98.8	98.8	95.5
Dead Nodes	1	5	1	1	4
Success Rate	99.4	57.8	96.66	94.81	81.37

Table 2. Evaluation of different routing algorithms for Network size 90 Nodes

The above table shows a comparison between different algorithms the constant number of nodes and packets generated. Each node consumes equal energy for the transmission purpose, viewing the results of the simulations; it is vivid that ELBGR algorithm performs much better than greedy algorithm as for the success rate, throughput, & number of alive nodes. The network lifetime using ELBGR is longer than the Greedy algorithm. ELBGR also outperforms EEAR, EAGR and HHEAA in terms of throughput and success rate yet the number of alive nodes after the simulation is similar to that of EEAR and EAGR. We performed so many simulations to attain the complete picture of the results of different algorithms. Our main objective is to evaluate different algorithms with changing network size and the traffic load, all other parameters remains constant in our simulations. The main aim of our simulation is to evaluate the performance of the proposed algorithm that what type of behavior it shows with increased number of the nodes and how it scales and performs better for different sizes of the networks. Viewing the detailed picture of simulation result, it is clearly seen that the delivered number of packets increases with increasing number of nodes of the network. It is also observed that the number of alive nodes is far greater using ELBGR routing algorithm than in number of alive nodes using greedy routing. This increases the life-time of the network.

F. RESULTS & PERFORMANCE COMPARISON

This Section describes the experimental results of the simulations & briefs the observations of the strengths and the limitations of the proposed algorithm ELBGR in comparison with the Greedy Routing algorithm, EEAR, EAGR and HHEAA to check the proposed scheme efficiency.

The results that are collected from the experiments include the comparison of the total number of packets delivered, the total number of packets dropped, the success rate, remaining live and dead nodes after simulation and throughput by different routing protocols. It is observed that the proposed algorithm utilizes maximum network resources and results far better than above mentioned existing routing protocols. It has been observed that in Greedy routing algorithm due to excessive use of few nodes those nodes deplete their energy very soon and results in the formation of holes, greedy algorithm don't consider the remaining energy of the nodes and the dead node simply drops the data packet. Hence using greedy algorithm so many packets are dropped and the data delivery rate is quite lower than ELBGR. The proposed algorithm ELBGR also performs better than EEAR, EAGR and HHEAA in terms of number of packets delivered and dropped (Figure 4, Figure 5).

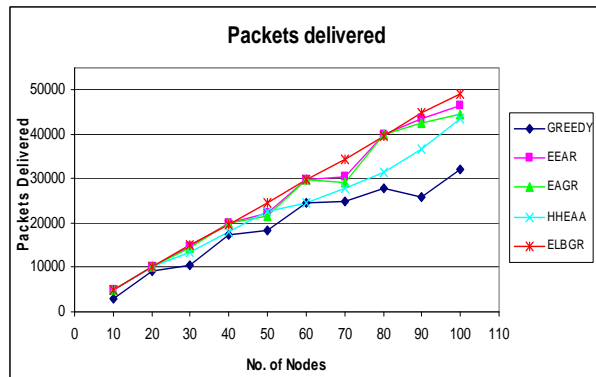


Figure 4. Comparison of No. of packets successfully delivered

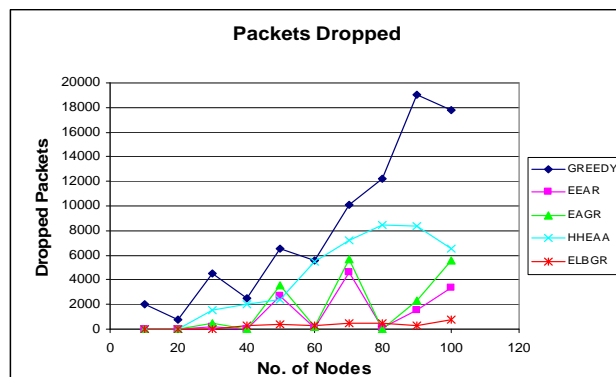


Figure 5. Comparison of No. of packets dropped

Our proposed algorithm ELBGR considers the current energy levels of nodes & don't sends the data towards the lower energy nodes that results in increased number

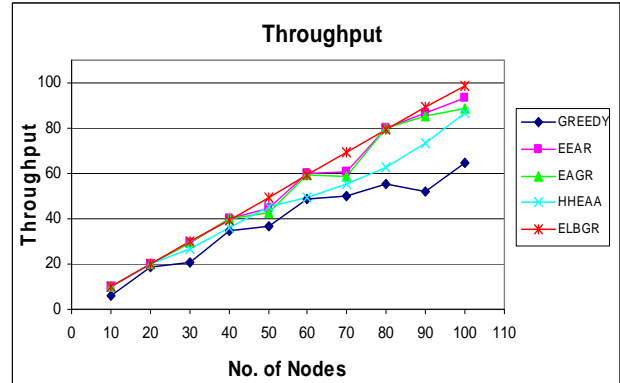


Figure 6. Comparison of Protocols Throughput

of packets delivery and very low number of packets are dropped using ELBGR. Hence it can be seen from Figure 6 that the throughput of the proposed algorithm is far better than Greedy algorithm as well as then EAGR, EEAR and HHEAA.

Figure 7 provides the comparison of the success rate of the Greedy algorithm, EAGR, EEAR and HHEAA and ELBGR; it is opaque that ELBGR performs very well.

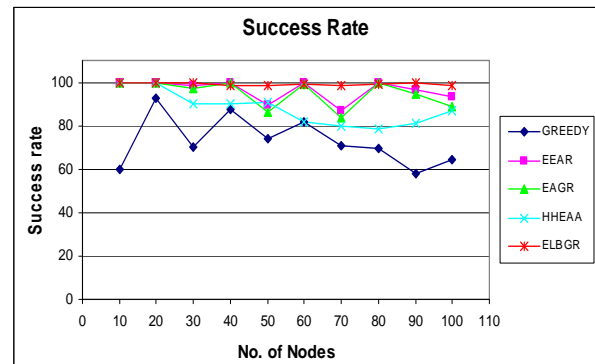


Figure 7. Comparison of Protocols Success Rate

Figure 8 shows that the number of alive nodes after the simulation duration using the proposed algorithm ELBGR is greater than the number of live nodes using Greedy algorithm and HHEAA. Number of alive and dead nodes using EAGR and EEAR are almost similar to that of ELBGR. (Figure. 8, Figure. 9).

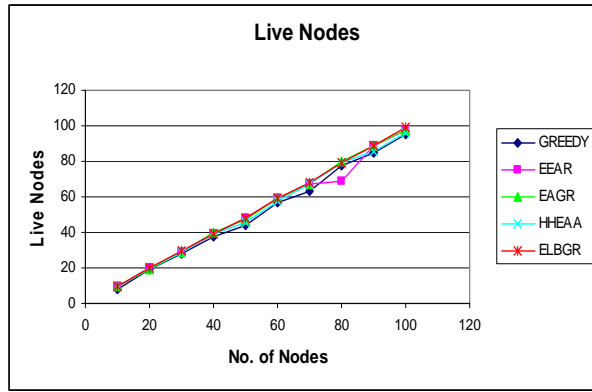


Figure 8. No. of Live Nodes Comparison

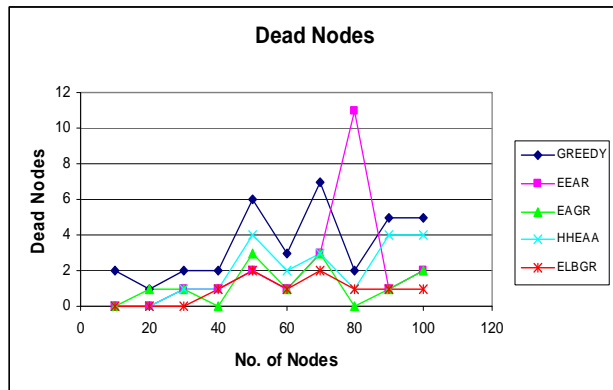


Figure 9. No. of Dead Nodes Comparison

Figure 10 shows a comparison of time delay in data delivery using both Greedy and ELBGR routing schemes, ELBGR take more time as it has to compute multiple parameters before forwarding the data.

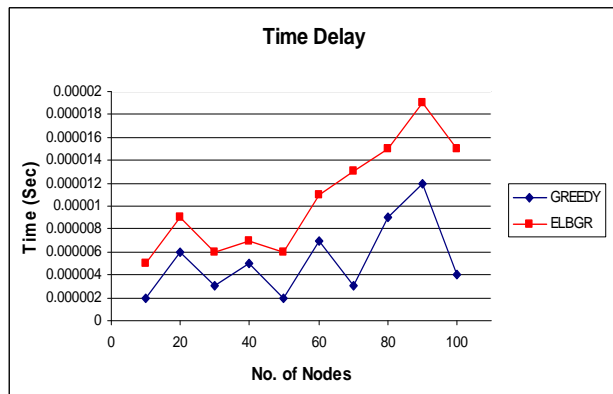


Figure 10. Time Delay Comparison

In the last the above mentioned algorithms are evaluated on the basis of their overall performance, which is measured by taking percentage of their average throughputs for different sizes of networks and different amount of traffic generated. It is clearly seen from Figure 11 that the performance/ efficiency of the proposed algorithm ELBGR (98.92%) is far better than

that of Greedy algorithm (70.56%). It can also be seen that ELBGR also outperforms EEAR, EAGR and HHEAA having average throughput percentage 95.4%, 93.52 and 84.76% respectively.

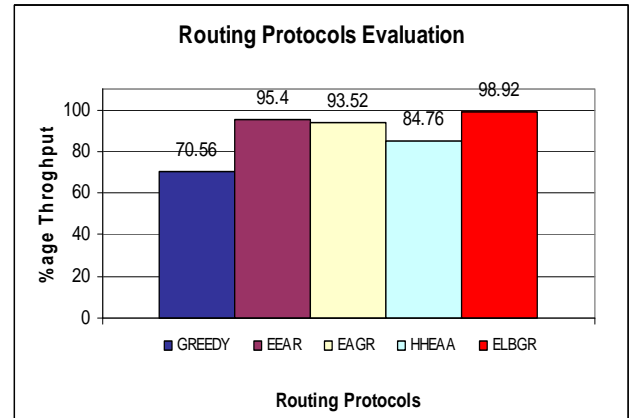


Figure 11. Throughput Percentage Comparison

G. DISCUSSION

Using the same input parameters, different experiments have been performed for different sized-network (having different amount of nodes). Same energy is consumed for each transmission by each node. It is observed that the proposed algorithm ELBGR outperforms Greedy routing algorithm in all aspects including successful packets delivery, throughput and remaining number of live nodes except the data delivery time-delay component, for greedy routing computation time is lesser than ELBGR as ELBGR has to compute multiple parameters. It is also seen that the ELBGR also outperform some pre-existing routing protocols (EEAR, EAGR and HHEAA) in the success rate and throughput.

A thorough picture of the detailed simulation results is presented using the graphs. Initially the Greedy routing algorithm is compared for different parameters with the ELBGR as our main focus is to make a comparison of the proposed algorithm's (ELBGR) performance with the Greedy routing strategy to evaluate the proposed algorithm's performance. Later on a comparison of ELBGR is also made for EEAR, EAGR, and HHEAA for the above parameters to check ELBGR efficiency. It can be concluded after vivid observation of the results extracted from the simulations that ELBGR is better, more efficient and reliable as compared to Greedy algorithm, EEAR, EAGR & HHEAA routing schemes and it increases the network life-time with maximum utilization of the resources for maximum possible duration.

VI. CONCLUSION AND FUTURE WORK

The main theme of this research to handle the wireless medium's broadcast nature along with energy constraints associated with the wireless sensor networks

that put the great effect on the network life time, scalability and reliability. In this research work an energy aware routing algorithm ELBGR (Energy aware and Load Balancing Geographic Routing) is proposed, implemented and evaluated. The main idea used to propose the routing scheme is to distribute the workload of the whole network among all nodes evenly so that the network resources may be utilized at their maximum capacity. The proposed routing scheme is compared mainly with the Greedy routing, the main limitation of greedy routing is that it heavily utilizes few nodes for data transfer that are at optimal distance from the destination and an uneven workload distribution among the nodes of the network has been observed. Therefore greedy routing algorithm results in the formation of many holes (over-utilized nodes), that blocks the further packets transmission and the network lifetime becomes very short even there are sufficient resources available in the network. These limitations of greedy routing are mainly focused in this research and the proposed algorithm very smartly removes these shortcomings. Initially the proposed algorithm tries its level best to avoid the creation of holes, but when all the neighbors utilizes almost their all resources and they reach at the lower threshold levels then the holes formation can't be avoided. So therefore when the holes are created, the proposed algorithm selects the remaining live nodes for data transmission rather than sending data to the dead end that may result in the decreased throughput. The proposed algorithm ELBGR works on the raw information that is just locally collected i.e. location of the nodes, nodes' energy levels and PRR values. This locally collected information is then exchanged among one-hop neighbors, every node maintains a local-table of its all one-hop neighbors, this local-table helps the node in making decision to either select any neighbor as forwarding candidate or not. The proposed scheme chooses multiple forwarding candidates initially based on remaining energy levels and the reception rates of the nodes and then they are prioritized based on their distance from the destination/target, using this strategy the successful transmission rate is significantly improved. In this research work, we have presented the principles of the local behavior of ELBGR, further more the efficiency of ELBGR has been concluded that the proposed routing strategy (ELBGR) is more efficient than Greedy routing strategy in all aspects (parameters that are evaluated) as well as the proposed algorithm ELBGR also out performs some other routing protocols (HHEAA, EAGR, and EEAR) in most of the aspects. The detailed analysis using graphs is describes in previous section.

The WSN field is in the evolutionary stage since last few decades, therefore having a great research potential for the researchers. For proving the proposed algorithm ELBGR more reliable and efficient, this algorithm may

be extended in some directions in future. For this research work, the static or immobile sensor nodes are considered, a single fixed destination is taken into consideration and the packets size is assumed fixed. For future research work the mobile sensor nodes may be considered with multiple destination nodes and variable size packets. Energy consumption can be calculated depending upon the radio ranges and the distances of nodes.

VII. BIBLIOGRAPHY

- [1] Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks", IEEE Communication Magazine, Aug. 2002.
- [2] A.Khetrapal, "Routing techniques for Mobile Ad Hoc Networks Classification and Qualitative/ Quantitative Analysis," Department of Computer Engineering, Delhi College of Engineering University.
- [3] J. Fraden. *A hand book of modern sensor: Physic, design, and application*. Birkauer, 2004.
- [4] J. N. Al-Karaki and A. E. Kamal. "Routing techniques in wireless sensor networks: a survey". In IEEE Wireless Communications, Volume 11, pp. 6- 28, 2004.
- [5] B. Krishnamachari, D. Estrin, S. Wicker, "Modeling data centric routing in wireless sensor networks", in: Proceedings of IEEE INFOCOM, New York, June 2002.
- [6] C. K. L. Lee, X. H. Lin, and Y. K. Kwok, "A Multipath Ad Hoc Routing Approach to Combat Wireless Link Insecurity", Proc. ICC 2003, vol. 1, pp. 448-452, May 2003.
- [7] G. Acs and L. Buttyabv. "A taxonomy of routing protocols for wireless sensor networks," *BUTE Telecommunication department*, Jan. 2007.
- [8] KARP B, KUNG H. GPSR: Greedy perimeter stateless routing for wireless networks [C]//Proceeding of the 6th Annual International Conference on Mobile Computing and networking, Aug 6-11, 2000, Boston, MA, USA. New York, NY, USA: ACM, 2000: 243-254
- [9] K. Sohrabi, J. Gao, V. Ailawadhi and G.J. Pottie, "Protocols for Self-Organization of a Wireless Sensor Network", *IEEE Journal of Personal Communications*, vol. 7, issue 5, pp. 16-27, Oct. 2000.
- [10] Y. XU, J. HEIDEMANN, AND D. ESTRIN, *Geography-informed energy conservation for Ad Hoc routing*, in MobiCom '01: Proceedings of the 7th annual international conference on Mobile computing and networking, New York, NY, USA, 2001, ACM, pp. 70-84.
- [11] F. Kuhn, R. Wattenhofer, and A. Zollinger, Worst-Case Optimal and Average-Case Efficient Geometric Ad-Hoc Routing, *ACM MOBIHOC* (2003).
- [12] Y. Yu, D. Estrin, and R. Govindan, (Geographical and Energy-Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks", UCLA Computer Science Department Technical Report, UCLA-CSD TR-01-0023, May 2001.
- [13] Shibo Wu et al, Power-aware single- and multipath geographic routing in sensor networks, Department of Computer Science, Fulton School of Engineering, Arizona State University, Tempe, AZ 85287, United States
- [14] J. Chen, R. Lin, Y. Li, Y. Sun, "LQER: A Link Quality Based Routing for Wireless Sensor Networks", *Sensors* 2008, 8, 1025-1038.
- [15] Razia Haider and M. Y. Javed, "EAGR: Energy Aware Greedy Routing in Sensor Networks", in International Conference on Future Generation Communication and Networks FGCN 2007, VOL 2, 344349-344349
- [16] Ifrah, F. K. and Javed, M. Y., "Hole Healing Energy Aware Algorithm for Wireless Sensor Networks", International

AUTHOR'S PROFILE

Journal of Computer Theory and Engineering (IJCTE), (ISSN: 1793-821X), Singapore

- [17] Munazza, Y., Abid, A. M., Javed, M.Y. and Atif, N., "EEAR: Efficient Energy Aware Routing in Wireless Sensor Networks", Proceedings of the 7th IEEE International Conference on ICT and Knowledge Engineering (ICTKE-2009), December 1- 2, 2009, (ISBN: 978-1-4244-4514-1), Bangkok,Thailand, pp. 57-62.
- [18] H. Hassanein, J. Luo, "Reliable Energy Aware Routing in Wireless Sensor Networks",dssns,pp.54-64, Second IEEE Workshop on Dependability and Security in Sensor Networks and Systems, 2006.
- [19] Archana Bharathidasan, Vijay Anand Sai Ponduru, "Sensor Networks: An Overview", Department of Computer Science, University of California, Davis, CA 95616.
- [20] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless sensor Networks", Elsevier Journal of Ad Hoc Networks. v3 i3. 325-349.
- [21] W. Heinzelman, A. Chandrakasan, H. Balakrishnan, "Energy-efficient communication protocol for wireless sensor networks", in: Proceeding of the Hawaii International Conference System Sciences, Hawaii, January 2000.
- [22] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. "Next Century Challenges: Scalable Coordination in Sensor Networks". In *MobiCom 99*, Seattle, USA, Aug. 99.
- [23] C. Chong and S. P. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges", in Proceedings of the IEEE, vol. 91, no. 8, Aug. 2003.
- [24] G. Acs and L. Buttyabv. "A taxonomy of routing protocols for wireless sensor networks," *BUTE Telecommunication department*, Jan. 2007.
- [25] E. Ilker Oyman and Cem Ersoy, "Overhead Energy Consideration for Efficient Routing in Wireless Sensor Networks", Computer Engineering Department, Bogazici University, Istanbul, Turkey.
- [26] F. L. Lewis, "Wireless Sensor Networks", in the Proceedings of Smart Environment Technologies, Protocols, and Applications, New York, 2004 (to appear).
- [27] Lakshmana Prasanth Nittala Venkata, "Routing Algorithms for Large Scale Wireless Sensor Networks", Master thesis for Computer Science, Texas A&M University, Dec 2004.



Nazia Perwaiz, Received Masters Degree (MS software Engineering) from National University of Science & Technology, Pakistan and a candidate for PhD studies in the field of sensor networks.



Professor Dr. Muhammad Younus Javed, department of computer engineering, NUST College of Electrical & Mechanical Engineering, National University of Sciences & Technology, Islamabad.

Custom Software under the Shade of Cloud Computing

Sheikh Muhammad Saqib¹, Muhammad Ahmad Jan¹, Bashir Ahmad¹, Shakeel Ahmad¹ and Muhammad Zubair Asghar¹

¹Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

Abstract— A thing which is fruitful for every one, who will not use that? Cloud computing is such shaded tree, where any one can sit. Each automated organization has its own custom software which is the replacement of manual work. Custom software is installed on each organization's machine and connected with database server. Server may be crash; simple fault in local network can occur or complexities about maintenance can hang whole custom application. So organizations can run their custom software in cloud computing environment and run their application on smooth way without any tension or confusion. Here author suggest a strategy how Desktop Based Custom Application (DBCS) can be converted in to Cloud Based Custom Software (CBCS). Since CBCS will be uploaded on FTP so organization's users can access outside the organization. Here authors also provide the concept and implementation of Tinny Application (TA), which will work as barriers. By using TA, organization can take advantages of cloud computing, but CBCS will not work outside organization as DBCS.

Keywords- Cloud Computing, Custom Software, CBCS, DBCS, TA, FTP

I. INTRODUCTION

Fundamental change in information technology that have emergence of phenomenon known as cloud computing. In cloud computing services are invented, developed, scaled, deployed updated and paid for. Applications containing such services are used for general purpose and provide wonderful economies of scale if they are consolidated supplied. A cloud provides resources on demand their charges depends on the use of services. These services may be dedicated services and based on large farms of inexpensive [5] [6].

Delivery of storage resources and computing to customer on demand is the major theme of cloud computing. It is excellent alternative for educational institution which has short budget. Such institutes can operate their systems efficiently by hiring some services for computers and network devices. Universities can take all benefits of available cloud's applications for enabling users or students to perform different tasks [7].

Any application based on cloud concept can get all advantages of cloud computing. Custom software for particular organization which is known as Desktop Based Custom Software (DBCS) can get all benefit of cloud computing and can run away from flaws associated with desktop application. CBCS will be a dedicated service for an organization and

stored on third party space and database. DBCS can be easily converted to Cloud Based Custom Software (CBCS). CBCS requires a same development framework as of DBCS; only difference is that CBCS can be uploaded on third party space. Here we are exploring DBCS which are developed in VB.Net framework. In such framework DBCS can be easily converted to CBCS with the help of ASP.net. Now this dedicated service (WebPages and database) can be easily uploaded on any third party space. Now this CBCS for particular organization can take all benefits of cloud computing. As users of CBCS can access this service out the organization, so here authors suggest a concept of Tinny Application (TA). TA will restrict the access of CBCS outside the organization.

II. CLOUD COMPUTING

Maintaining data and application through central remote server and internet is known as cloud computing. Without installation, consumers and businesses can use applications and can access their private files on any compute any where with internet. His technology allows for much more efficient computing by centralizing storage, memory, processing and bandwidth. Yahoo, Gmail, Picturtrail etc are the examples of cloud computing [3]. Utilizing the hosted services over internet comes under the umbrella of cloud computing. It means we can work on hosted document, different applications that can look like a services and can be stored on any service provider from any place. Cloud computing includes different services such as SaaS (Software as a service): with the help of front end tool, users can interact with it, PaaS (Platform as a service): by using infrastructure of providers users can create their applications, IaaS (Infrastructure as a service) it provides virtual server and memory. Applications based on cloud computing are usually optimized, simple and easy to use. It reduces learning curve which is required for new staff and increases efficiency and communication capacity between packages of various software [2][4].

III. CUSTOM SOFTWARE

The main intention of custom software, bespoke software, is to fulfill the goals of an organization. The ultimate responsibility of its development lies either on the shoulders of a software development group or independent developer. And

it is there to fulfill all possible preferences and expectations of intended audience. Custom software exists to serve companies ranging from small size through medium size and ultimately by large organizations for their core and vital functions.

Large companies commonly use custom software for critical functions such as content management, inventory management, customer relationship management and human resource management [1]. Mostly these applications are desktop based. Successfully installation of custom software requires different equipments such as:

Switches: Use for connecting all clients (4 port switch, 8 port switch, 16 ports switch, 24 ports switch etc).

Database Server: A System for storage purpose is required (RAID or Core I3 or Core I7 system, database software, operating system, ant viruses).

Backup Devices: External storage devices such as compact disk, external hard disks, tap devices; logical backup devices etc are required.

Compatible framework i.e. (Microsoft .NET Framework v3.0 or higher) must be installed on the client's machine.

A database administrator is required for backup & recovery and others database issues.

Successful reinstallation is needed after crash of server.

IV. REPLACING DBCS TO CBCS

Custom software for any organization is just like a desktop application which is developed in VB.Net. For getting different advantages of cloud computing we can easily convert such software as a service. VB.Net supportable features also merged in ASP.Net such as textboxes, labels, button, grids, data storage and even reports. ASP.Net web pages can be remotely accessed. We can easily convert whole desktop custom software into ASP web pages (CBCS). ASP web pages can be easily uploaded using FTP. There are many companies which will provide the FTP spaces and database for storage purpose. Now through single URL, all clients can access CBCS. And such work will be done through web browser instead of desktop based forms. All clients in an organization can use CBCS only by connecting their machine with internet. Each user can login through password. Fig-1 shows that desktop based form is converting into web based form.

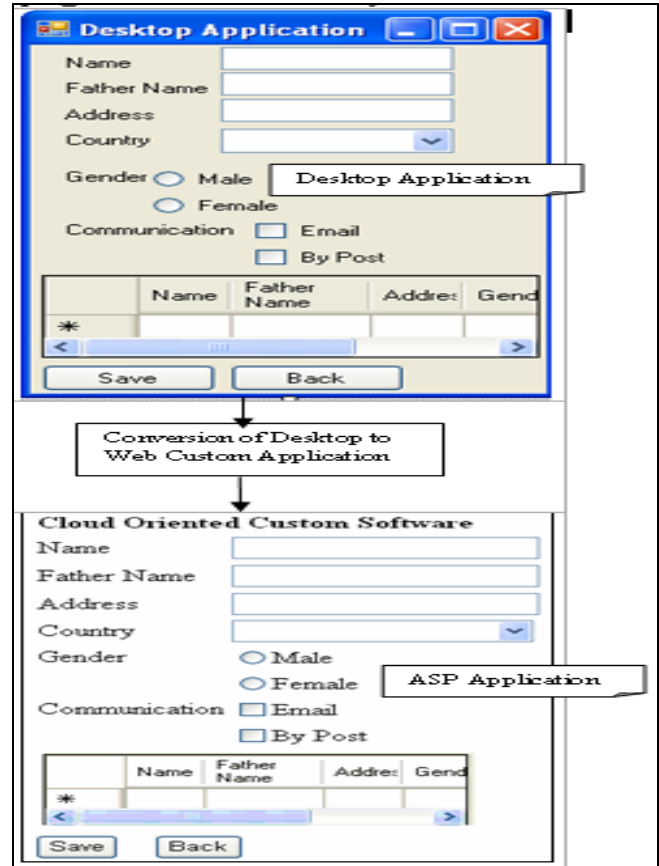


Figure 1. Replacing Desktop form to Cloud Based Form

V. WHY WE WANT TO REPLACE DBCS TO CBCS

CBCS does not require all requirements which are necessary for DBCS as mentioned in above such as (database server, backup devices). But in CBCS some switches can be used for providing and expanding internet. With the help of CBCS there is no need of daily backups, no fear of lights off, if client machine is affected then CBCS will require only reinstallation of operating system. There will also no need to install any compatible frame work for customized software. CBCS require annual cost for FTP and database storage while DBCS requires maintenance cost. We can not say in this cost DBCS is better, because besides maintenance cost, maintenance procedures is also some time becomes more complex, technical and over headed.

VI. MAJOR DRAWBACK OF CBCS

Since CBCS is only for a particular organization so it should not be accessed outside the organization i.e. if software for library is made up of CBCS then relevant user can issue the book to any person at his home. Although it is work or responsibility of that user but issuing a book at home leads to mismanagement. Book must be issued at library, when issuer has a book in his hand. Besides this issue CBCS can get all benefits of cloud computing. Such type of matter can not be occurred in DBCS because desktop application is install only on client machine at organization or office, so users can not access the application outside organization.

VII. SOLUTION OF DRAWBACK OF CBCS USING TA

CBCS should not be accessed at outside the organization, this issue must be handled by the organization authority not any third party member. So it can be done by using password handler application. Tinny Application (TA) a small application which will handle the password policy of CBCS. CBCS will be uploaded on ftp while TA will be installed on each machine in relevant organization. Every one in an organization can work on CBCS because TA will be installed over here while at out side organization CBCS can not be worked because TA will not be installed at out side the organization. Following figure is showing that organization's clients can work on CBCS by using TA while clients other than organization can not work on such type of CBCS.

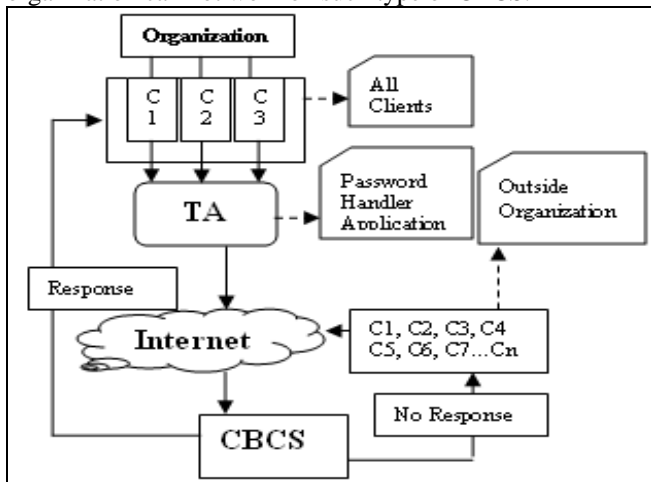


Figure 2. Accessing CBCS using TA

VIII. IMPLEMENTATION OF TA

Since each user in an organization have his own password, by using such password he can login on CBCS. Login interface of CBCS will have username, original password and provisional password. Original password will be the permanent password and provisional password will be randomly generated through TA with the entry of original password. Then user can copy this random provisional password from TA and paste it on

login interface of CBCS. These provisional passwords will be deleted after login on CBCS. If user gets provisional password from TA and can not be login on CBCS then, this provisional password will be removed from database after 30 seconds. Any one (even organization member who have their original password) out side the organization can not be login because they can not get random provisional password. Tinny application will also be connected with hosted database so there will be no need of local database.

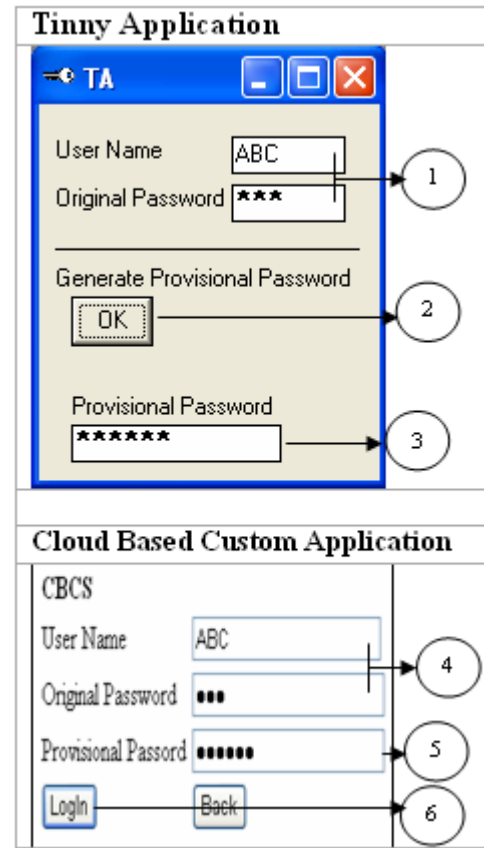


Figure 3. Working of TA

Steps shown in figure will be described as

1. Organization Member will enter user name and original password.
2. After Step-1, user will click on 'OK' button.
3. After Step-2, random provisional password will be generated in third textbox and stored in hosted database. Now user will copy this provisional password and past on provisional password text box of CBCS.
4. Now on CBCS, organization member will enter the user name, original password and
5. paste provisional password in next text box.
6. User can click on login button for work on CBCS.
7. When user name, original password and provisional password will be compared with stored entries, if all are matched then user can work on CBCS.
8. After login, provisional password will be deleted from hosted database.

IX. RESULTS AND CONCLUSION

Technology expert assumed that in 2020, every thing will be worked on cloud environment through cyberspace based applications on network devices [8]. This time is like a road where every one is going to 2020 with taking the luggage of cloud computing. Now it is time for automating each organization under the shade of cloud computing. CBCS concepts will move the organization to environment and they will leave the DBCS at their organization due flaws associated with DBCS. CBCS will have many advantages over DBCS.

TABLE I. ADVANTAGES OF CBCS OVER DBCS

<i>Comparison of CBCS and DBCS</i>	
DBCS	<i>Development for DBCS is mandatory.</i>
CBCS	<i>Development for CBCS is also mandatory.</i>
DBCS	<i>Database administrator for DBCS must be technically sound. He must have knowledge about all securities issues about DBMS.</i>
CBCS	<i>Database administrator for CBCS just has knowledge about setting permissions.</i>
DBCS	<i>Database software is required i.e. DBMS.</i>
CBCS	<i>No need of Database software.</i>
DBCS	<i>Devices for Backup is required i.e. compact disk, external storage devices, flash sticks etc.</i>
CBCS	<i>No need of Backup devices.</i>
DBCS	<i>Efforts on backup and recovery is required i.e. setting time, if backup is generated automatically. Light must be available on time of backup. After taking the backup, backup devices must be kept in safe and secure places.</i>
CBCS	<i>No efforts require on backup and recovery</i>

DBCS	<i>After system crash, efforts on complete installation of operating system & DBMS, proper handling of personal database creation of supportable routines for custom software i.e. DSN etc will be required.</i>
CBCS	<i>After system crash, only installation of operating system will required for connecting with internet.</i>
DBCS	<i>A new system can not be easily replaced.</i>
CBCS	<i>A new system can be easily replaced at once and CBCS will work properly on spot.</i>
DBCS	<i>DBCS requires its supportable framework.</i>
CBCS	<i>CBCS does not require any supportable frame work, just internet connections.</i>
DBCS	<i>Database server is required (Core I3, Core I7 etc).</i>
CBCS	<i>No need of database server.</i>
DBCS	<i>DBCS requires technical securities on local area network i.e. network cable should not be crossed over sensitive area with respect to electricity, entrance place etc. because fault in single place of local area network can stop the DBCS.</i>
CBCS	<i>No need of technical securities on local area network even without local area network, CBCS can work properly with the connection of internet.</i>

Since custom software is installed on organization's computer, so they can access only on those computers not outside the organization. But CBCS are web based, so they can be accessed outside the organization with its own user's authentication. This may leads to security issues. Here author gives a concept of TA which will allow CBCS to be accessed over the organization's computers and restrict the outside the accession of CBCS.

REFERENCES

- [1] Custom software,
http://en.wikipedia.org/wiki/Custom_software
- [2] Carnegie Mellon M.S , The Basic Concepts of Cloud Computing, <http://www.articlesbase.com/databases-articles/the-basic-concepts-of-cloud-computing-1284980.html>, 2009
- [3] cloud computing,
http://www.wikinest.com/concept/Cloud_Computing
- [4] Sathishkumar, Cloud Computing: Best Practices,
<http://www.tech2date.com/cloud-computing-best-practices.html>, 2011
- [5] Sean Marston, Cloud computing — The business perspective, *Decision Support Systems* 51 (2011), P. 176-189, ELSEVIER.
- [6] Arnon Rosenthal, Cloud computing: A new business paradigm for biomedical information sharing, *Journal of Biomedical Informatics* 43 (2010) 342–353, ELSEVIER.
- [7] Tuncay Ercan, Effective use of cloud computing in educational institutions, *Procedia Social and Behavioral Sciences* 2 (2010) 938–942, ELSEVIER.
- [8] Janna Quitney Anderson, The future of cloud computing, June 11, 2010, <http://pewinternet.org/Reports/2010/The-future-of-cloud-computing.aspx>

About Author

SHEIKH MUHAMMAD SAQIB is with Gomal University, D.I.Khan, Pakistan. (E-mail: saqibsheikh4@hotmail.com). His research interest is in RUP, SOA and Cloud Computing. Some of his papers were presented in some conference such as 9th National Research Conference at SZABIST – Islamabad PAKISTAN, 2009, ISCS (International Symposium on Computer Science) at Namal College Mianwali PAKISTAN an Associate College of The University of Bradford, UK, 2011 and Ten (10) are published in international journals.

Lt. Dr. S Santhosh Baboo, Reader
P.G. Research Dept of Com. Science
D G Vaishnav College,
Arumbakkam, Chennai – 106.

V J Chakravarthy
Research Scholar
Dravidian University
chakku_vjc@yahoo.co.in

ABSTRACT

In this paper, we have proposed a new RAOA protocol (Right Angle or Ant Search). It is a reactive (on-demand) routing protocol. This is done in route reply (RREP) method. During route reply when more than one route replies are about to reach to source there are high chances that they create congestion at the last point. This congestion is a cause for possible collisions. We tried to reduce this possibility of collisions. We propose to include delay in RREP when RREP is one hop away from the source. We compared the performance of the our proposed protocol RAOA with three prominent routing protocols for mobile ad hoc networks, namely Ad hoc On Demand Distance vector (AODV), Ad hoc On-demand Multipath Distance Vector (AOMDV) and Destination Sequenced Distance Vector (DSDV). We have chosen four performance metrics, such as Average Delay, Packet Delivery Ratio, Routing Load, and Throughput. The performance simulations are carried out on NS-2. The performance differentials are analyzed using varying network size and simulation times. The simulation result confirms that RAOA performs well in terms of Average Delay, Packet Delivery Ratio, Routing Load, and Throughput.

KEYWORDS

Ad-hoc Networks, Collision, Congestion, Average Delay, Performance Analysis, Routing protocols, Simulation.

INTRODUCTION

There has been rapidly increase in the number of users of the wireless communications, from the satellite communication to the home wireless personal area network, over the last few years. Wireless communication has gained such popularity because of the ability of the wireless node to communicate with the rest of the world while being mobile. But the presence of a fixed

supporting structure limits the adaptability of wireless systems. Wireless networks can generally be classified as wireless fixed networks and wireless ad-hoc networks. Wireless LANs and cellular network can be considered as infrastructure aided wireless fixed network. Mobile Ad-hoc network can be considered as a special type of wireless ad hoc network formed without any infrastructure or any standard services. The multi-hop support in ad-hoc networks make it possible to communicate between nodes outside direct radio range of one another, which makes it different from wireless LANs.

INTRODUCTION TO MOBILE AD-HOC NETWORKS (MANET)

A mobile ad-hoc network (MANET) is a collection of wireless nodes that can dynamically be set up anywhere and anytime without using any pre-existing network infrastructure. It is defined as an autonomous system of mobiles nodes. Mobile hosts connected to wireless links are free to move randomly and often act as routers at the same time. Mobile Ad-hoc networks are emerging as the next generation of networks. Mobiles nodes are capable of transmitting the packets to the nodes which are in proximity. If a mobile node has packet to send to other mobile nodes, which are out of its range, then the nodes within its range forwards packets to the next node(s) until packet reaches the specified destination. This is why MANETs are also called mobile multihop wireless networks. MANETs can be setup between few nodes or can be extended by connecting to fixed network. The system may operate in isolation, or interface with a fixed network. MANET nodes are equipped with wireless transmitters and receivers using antennas which may be omni directional (broadcast), highly-directional (point-to-point), or some combination thereof. At a given point in time, depending on the nodes positions and

their transmitter and receiver coverage patterns, transmission power levels and co-channel interference levels, a wireless connectivity in the form of a random, multihop graph or Ad-hoc network exists between the nodes. This Ad-hoc topology may change with time as the nodes move or adjust their transmission and reception parameters. The nodes may be located in or airplanes, ships, trucks, cars, perhaps even on people or very small devices, and there may be multiple hosts per router.

OBJECTIVE & OVERVIEW OF THE PROPOSED PROTOCOL

A. Objectives

In this paper, we propose to design a Multi-Path Routing Protocol, which sends the packets in alternative path, which has the following objectives:

Initially, we present a high efficient solution that seeks to utilize idle or under-loaded nodes to reduce the effects of congestion. To work out this, we highly enhanced the geographical routing to allow a source to select different paths to make the packet to reach the destination. First, we propose multi-path solutions for geographic routing which has less effective results, at the end, we likely to propose right angled biased geographical routing technique (RAOA), a lightweight, stateless, Geographical forwarding algorithm, as cost-effective complement to greedy routing. The above RAOA protocol routes packets in straight path i.e. 90° from the source, instead the shortest path, towards the destination.

To reduce the congestion during transmission of packets; we propose two more congestion control mechanisms that highly enhance RAOA protocol.

Biased Node Packet Scatter (BNPS) is a very light weight method mechanism that partially aims to transient congestion by locally splitting the traffic along multiple paths to avoid congested hotspots.

Node-to-Node Packet Scatter (NNPS) is also a mechanism but aim to transmit packets to longer term congestion, when BPNS fails.

The performance of the above two mechanism had been evaluated in term RAOA by using a high-level simulator, a packet-level simulator (NS-2). The results show that RABGR is a practical and efficient multipath routing algorithm. We have evaluated BNPS and NNPS using NS2.

2. Right Angled Biased Geographical Routing or ANT SEARCH (RAOA)

The requirements of the RAOA algorithm are as follows. In addition, we present simulation results that show that BGR achieves good performance with a low overhead.

Design goals

Wireless network with coordinate based routing. To have sensor networks, we require stringent energy and computational constraints, which characterize these networks.

The requirements of the geographic routing protocol:

1. **Low communication overhead** – packets sent by the sensor nodes are very small e.g. the maximum packet size is 29 bytes.
2. **Simplicity** – The routing algorithm must have low computational overhead e.g. 4 kB of RAM.
3. **Low state** – nodes much maintains a minimal amount of state i.e. no per-flow or per-path state in network. In addition, to avoid the hotspots in the considered wireless networks, a multi-path algorithm should be there, that must be able to provide a large number of path i.e., 90° , with few common hops without increasing routing failures, as compared to the single-path greedy routing.

Explanation of the Right Angled Biased Geographical Routing (RAOA)

The main idea in our solution is to reduce the congestion during the transmission of packets from source to destination, for that we inserted a “**BIAS**” i.e. the angle in each packet, which determines the straight line path from the source so that the packets move towards the destination. Here the term bias is a measure angle of which the packets take from the source from greedy route and also indicates the side of deviation. In our discussion, the term bias is treated at each hop as an angle i.e., 90° . Instead of routing greedily towards the destination. Our proposed protocol “RAOA” routes greedily towards the point P2 (target point) situated at a predefined distance from the current node point P1 such that the angle between the lines P1 and P2 is equal to the bias i.e. angle 90° and finally the P3 node receives the packets. If the sending node doesn't find any node at 90° , instantly it will search (Ant Search) for the node which is very near to that node. Once it finds the very nearest node, it will send the packets continuously. Then that node finds the other node at 90° and sends the packets.

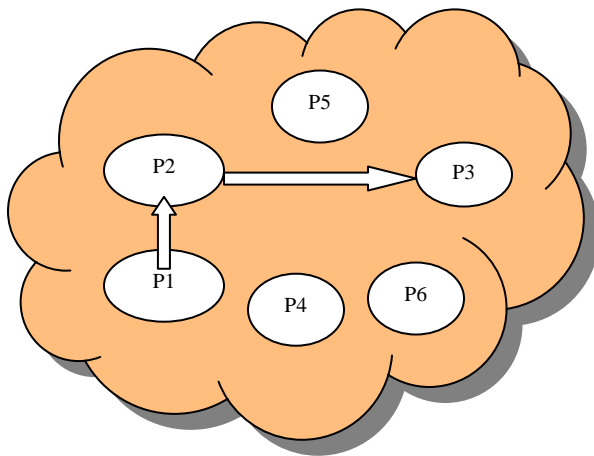


Figure 1: RAOA Forwarding

Minimising Congestion in Wireless Networks

In wireless networks, Congestion occurs when the wireless area around them is busy.

With networks congestion is mostly situated at the border of the network, with point to point communication congestion usually builds in the center. So avoid the congestion in the wireless networks, the way should be followed, i.e., we allow packets to route on alternate paths. This type of routing avoid the congestion is busy area in the wireless networks.

BNPS – Biased Node Packet Scatter

BNPS splits flows close to the congestion point. Each node monitors the congested status of all its neighbours and splits the flows that are going towards a congested neighbour, if the node itself is congestion. The scattered packets contain bias of 90° , such that the modified paths quickly move away from the original path.

NNPS – Node – to – Node Packet Scatter

If BPNS cannot successfully support the aggregate traffic, it will only scatter packets to a wider area or amplifying the effects of congestion collapse due to its longer paths.

Evaluation of BNPS and NNPS

In this section we present simulation results obtained through NS-2 simulations. We use three main metrics for our measurements: throughput increase, packet delivery ratio and delay among flow.

We ran tests on a network of 20 nodes, distributed uniformly on a grid in a square area of $1000\text{m} \times 1000\text{m}$. We assume events occur uniformly at random in a geographical area; the node closest to the event triggers a communication burst to a uniformly selected destination. To emulate this model we select a one set of random source-destination pair and run 20 second synchronous communications among the selected pair. The data we present is averaged over hundreds of such iterations. The parameters are summarized in Table 1.

Table 1.**SUMMARY OF PARAMETERS**

Parameter	Value	Parameter	Value
Number of Nodes	20	Link Layer Transmission Rate	2 Mbps
Area Size	1000m x 1000m	RTS / CTS	No
MAC	802.11	Retransmission Count (ARQ)	No
Radio Range	100m	Interface Queue	No
Contention Range	250 m	Packet Size	100B
Average Node Degree	90	Packet Frequency	40/s

SIMULATION MODEL

In this section, the network simulation is implemented using NS-2. The Network Simulator NS-2 is a discrete event simulator, which means it simulates such events as sending, receiving forwarding and dropping packets. For simulation Scenario and network topology creation its used OTCL (Object Tool Command Language). To create new objects, protocols and routing algorithm or to modify then in NS-2, C++ source code has been changed. The simulator supports wired and wireless and satellite networks. The simulations were conducted on Celeron processor at speed 1.0 GHz, 256 MB RAM running Linux.

PERFORMANCE METRICS

While comparing our proposed protocol with other three protocols, we focused on four performance measurements such as Average Delay, Packet Loss and Through Put.

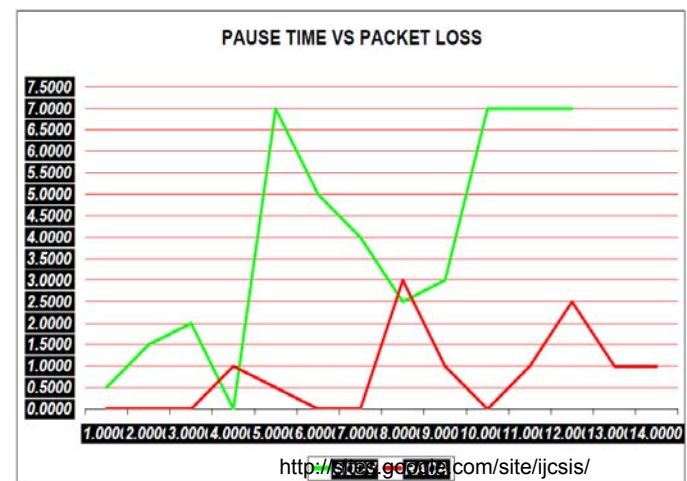
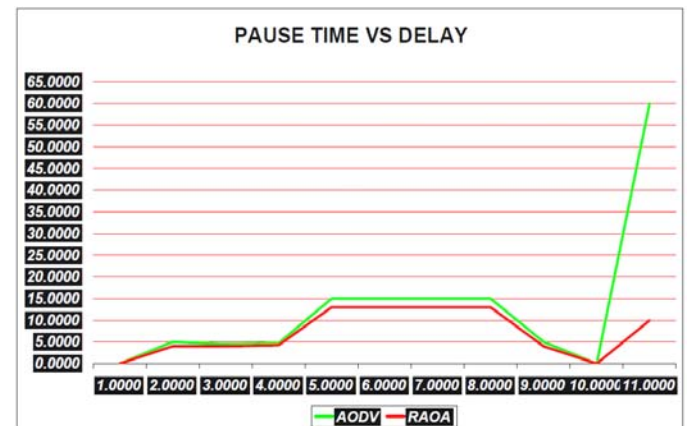
i. Average End to End Delay of Data Packets:

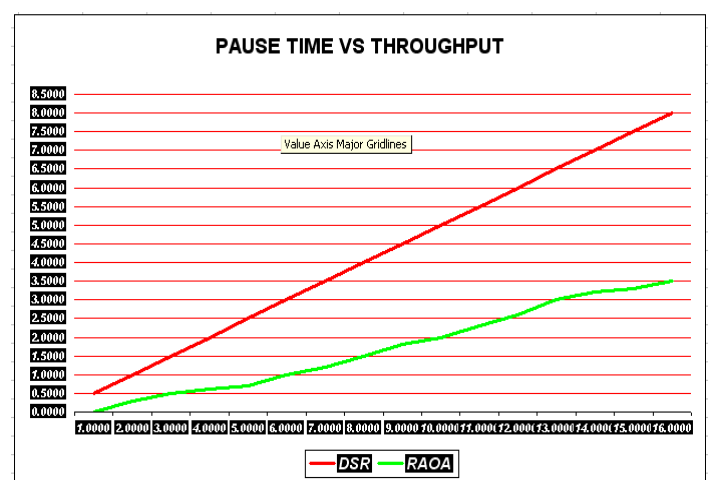
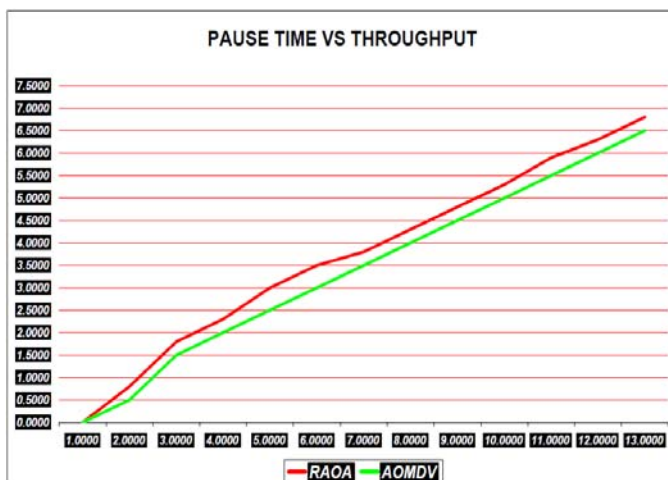
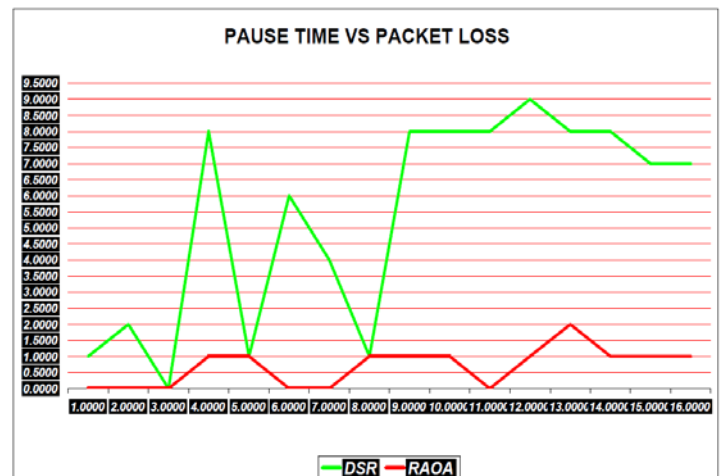
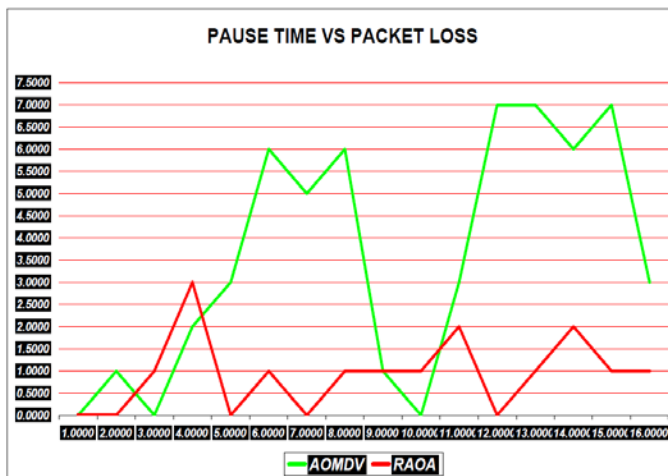
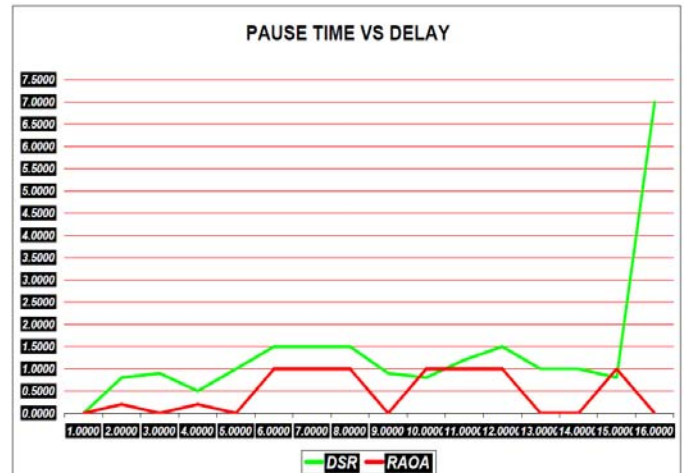
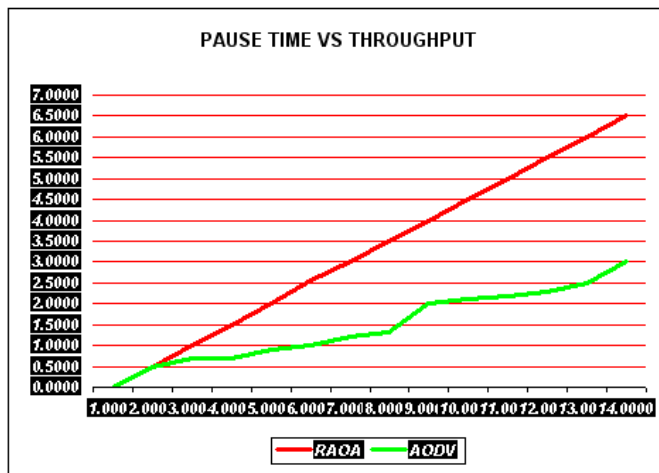
The average time from the beginning of a packet transmission at a source node until packet delivery to a destination. This includes delays caused by buffering of data packets during rout discovery, queuing at the interface queue, retransmission delays at a MAC, and propagation and transfer times. Calculate the

send (S) time (t) and Receive (R) Time (T) and average it.

ii. Packet Loss: It is a measure of the number of packets by the routers due to various reasons. The reason we have considered for evaluation are Collisions, Time outs, Looping, Errors.

iii. Throughput: It is the number of packets received successfully. In communication networks, such as Ethernet or packet radio, throughput or network throughput is the average rate of successful message delivery over a communication channel. This data may be delivered over a physical or logical link, or pass through a certain network node. The throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second or data packets per time slot.

Simulation Results



Conclusion

In this paper, we have developed a multipath routing protocol which attains confidentiality of packets in both routing and link layers of MANETs. This paper does the realistic comparison of three protocols namely AODV, AOMDV and DSR with our newly proposed Reactive (on-demand) multipath routing protocol RAOA. The significant observation is, simulation results agree with expected results based on theoretical analysis. As we expected, our routing protocol RAOA performance is the best considering its ability to maintain connection by periodic exchange of information, which is required for TCP, based traffic. As we know, routing protocol in grid environment is a rather hot concept in computer communications. In our future work we would be try to focus more on security issues.

REFERENCES

1. Liu, Kejun Deng, Jing Varshney, Pramod K, Balakrishnan and Kashyap "An Acknowledgment-based Approach for the detection of Routing Misbehavior in MANETs", IEEE Transactions on Mobile Computing, May 2007.
2. Based Approach for the Detection of Routing Misbehavior in MANETs", IEEE Transactions on mobile Computing, May 2007.
3. Anand Patwardhan, Jim Parker, Anupam Joshi, Michaela Lorga and Tom Karygiannis "Secure Routing and Intrusion Detection in Ad Hoc Networks", Third IEEE International Conference on Pervasive Computing and Communications, March 2005.
4. A. Patwardhan, J. Parker, M. Lorga, A. Joshi, T. Karygiannis and Y. Yesha "Threshold-based Intrusion Detection in Ad hoc Networks and Secure AODV" Elsevier Science Publishers B.V, Ad Hoc Networks Journal (ADHOCNET), June 2008.
5. Muhammad Mahmudul Islam, Ronald Pose and Carlo Kopp, "A Link Layer Security Protocol for Suburban Ad-Hoc Networks", in proceedings of Australian Telecommunication Networks and Applications Conference, December 2004.
6. Shiqun Li, Tieyan Li, Xinkai Wang, Jianying Zhou and Kefei Chen, "Efficient Link Security Scheme for Wireless Sensor Networks", Journal of Information And Computational Science, Vol.4, No.2, pp.553-567, June 2007.
7. S. Schmidt, H. Krah, S. Fischer, and D. Watjen, "A Security Architecture for Mobile Wireless Sensor Networks", In proceedings of First European Workshop on security in Ad-Hoc and Sensor Networks (ESAS 2004), August 2004.
8. A. Rajaram and Dr. Palaniswami "Detecting Malicious Node in MANET Using Trust Based-Layer Security Protocol", International Journal of Computer Science and Information Technologies, Vol. 1 (2), 2010, 130-137.
9. Ajay Jangra, Nitin Goel, Priyanka & Komal Bhatia "Security Aspects in Mobile Ad Hoc Networks (MANETs): A Big Picture", International Journal of Electronics Engineering, 2(1), 2010, pp. 189-196.
10. Abhay Kumar Rai, Rajiv Ranjan Tewari, Saurabh Kant Upadhyay "Different Types of Attack on Integrated MANET – Internet Communication".
11. Dr. S. Santhosh Baboo, B. Narasimhan "A Hop-by-Hop Congestion – Aware Routing Protocol for Heterogeneous Mobile Ad-hoc Network", International Journal of Computer Science and Information Security, Vol. 3, No. 1, 2009.
12. Marina and Das, "On-demand multipath distance vector routing in ad hoc networks", Ninth International Conference on Network protocols, pp.14-23, 2001.
13. Xiaoqin Chen, Haley M. Jones and Jayalath, "Congestion-Aware Routing protocol for Mobile Ad Hoc Networks", IEEE 66th conference on Vehicle Technology, pp.21-25, October 2005.
14. N. Vetrivelan, Dr. A V Reddy, "Performance Analysis of Three Routing Protocols for Varying MANET Size", Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol II, IMECS 2008 19-21, March 2008, Hong Kong.
15. Eric Setton, Xiaoqing Zhu and Bernd Girod "Congestion-Optimized Multipath Streaming of Video over Ad Hoc Wireless Networks".

16. S R Biradar, Subir Kumar Sarkar, Rajanna K M, Puttamadappa C “Analysis QOS Parameters for MANETs Routing Protocols”, International Journal on Computer Science and Engineering, Vol. 02, No. 03, 2010, 593-599.
17. P Chenna Reddy, Dr. P. Chandrasekhar Reddy, “Performance Analysis of Ad Hoc Network Routing Protocols”, Academic Open Internet Journal, ISSN 1311-4360, Vol-17, 2006.
18. Lucian Popa, Costin Raiciu, Ion Stoica, David s. Rosenblum, “Reducing Congestion Effects in Wireless Networks by Multipath Routing”.
19. Josh Broch, David A. Maltz, David B. Johnson Yih-Chun Hu, Jorjeta Jetcheva, “A Performance Comparision of Multi-Hop Wireless Ad Hoc Network Routing Protocols.
20. Rahul C. Shan, Adam Wolisz, Jan M. Rabacy, “On the performance of geographical routing in the presence of localization errors”.

AUTHORS PROFILE



Lt. Dr. S .Santhosh Baboo, aged forty, has around Twenty years of postgraduate teaching experience in Computer Science, which includes Six years of administrative experience. He is a member, board of studies, in several autonomous colleges, and designs the curriculum of undergraduate and postgraduate programmes. He is a consultant for starting new courses, setting up computer labs, and recruiting lecturers for many colleges. Equipped with a Masters degree in Computer Science and a Doctorate in Computer Science, he is a visiting faculty to IT companies. It is customary to see him at several National /international conferences and training programmes

both as a participant and as a resource person. He has Doctorate in Computer Science, he is a visiting faculty to IT companies. It is customary to see been keenly involved in organizing training programmes for students and faculty members. His good rapport with the IT companies has been instrumental in on /off campus interviews, and has helped the post graduate students to get real time projects. He has also guided many such live projects. Lt.Dr. Santhosh Baboo has authored a commendable number of research papers in international/national Conference/journals and also guides research scholars in Computer Science. Currently he is Reader in the Postgraduate and Research department of Computer Science at Dwaraka Doss Goverdhan Doss Vaishnav College (accredited at ‘A’ grade by NAAC), one of the premier institutions in Chennai.



V J Chakravarthy, done his Under-Graduation in Madras University and Post-Graduation in Bharathidasan University and Master of Philosophy Degree in Periyar University. He is currently pursuing his Ph.D in Computer Science in Dravidian University, Kuppam, Andhra Pradesh. Also, he is working as a Assistant Professor in P G Department of Computer Science, in the New College, (accredited at ‘A’ grade by NAAC), one of the premier institutions in Chennai.. He is having more than two years of research experience and seven years of teaching experience. His research interest includes Wireless - Ad-HocNetworks - Congestion reduction – Shortest Path detection, new protocols findings.

MULTIMEDIA DESIGN ISSUES FOR INTERNET TELEPHONY PROTOCOLS IN CURRENT HIGH PERFORMANCE NETWORKS

A.Jayachandran
Asst Prof,CSE Dept
PSN College of Engineering
and technology
Tirunelveli,INDIA
jaya1jaya1@gmail.com

Dr.R.Dhanasekaran,Principal
Syed Ammal Engineering
college,Ramnad,INDIA
rdhanashekar@yahoo.com

P. Rajan, Professor in MCA
dept, PSN College of
Engineering and Technology,
Tirunelveli,INDIA
prajan1968@gmail.com

ABSTRACT:

The rapid growth of media content distribution[5] on the internet in the past few years has brought with it commensurate increases in the cost of distributing the content. Underlying the internet access trends at a global scheme i.e. how people use the internet is a challenging problem i.e. typically addressed by network frames. We seek to understand the intrinsic reasons for the well known phenomenon of heavy tailed degree in the internet. AS graph and anywhere that the contrast to traditional model based on preferential attachment and centralized optimization the present degree of the internet can be explained by the evolution of wealth associated with each ISP. Our extensive numerical and simulation examples have shown that the proposed scheme achieves satisfied accuracy and computational efficiency. The effectiveness of the proposed detection and trace back methods are verified through extensive simulations and internet datasets.

Key words : Peer to peer , Network security, Security policy, Content delivery, 3G/4G networks, Grid and Cloud computing

1. INTRODUCTION

We describe how a scalable system of dense Wi-Fi sensors[6,7] can be built inexpensively. We build such a system and evaluate its performance. We provide specific example of why standard authentication and encryption scheme are inadequate to secure corporate Wi-Fi networks which motivate our solutions based on continuous monitoring of Wi-Fi networks. We show that, to provide comprehensive coverage for detecting security branches a dense deployment of RF sensor is necessary.

Magnetic and electromagnetic sensors do not require direct physical content and are useful for detecting proximity effects. Magneto resistive effect is a related phenomenon depending on the fact that the conductivity varies as the square of the applied flux density. Magnetic field sensors can be used to detect the remote presence of metallic objects. Eddy current sensors use magnetic probe coils to detect in the metallic structure such as pipes.

Thermal sensors are a family of sensors used to measure temperature or heat flux. Most biological organisms have developed sophisticated temperature sensing systems. Thermo resistive effects are based on the

fact that the resistance R changes with temperature T . For moderate changes the relation is approximately given for many metals by $\Delta R/R = \alpha R \Delta T$ with αR the temperature coefficient of resistance. The relationship for silicon is more complicated but it is well understood. Hence silicon is useful for detecting temperature changes.

Resonant temperature sensors rely on the fact that single crystal SiO_2 exhibit a change in resonant frequency depending on temperature change. Since this is a frequency effect it is more accurate than amplitude change effect and has extreme sensitivity and accuracy for small temperature changes.

2. On demand multimedia content

For a large content delivery networks that consists of hundreds are even thousands of geographically distributed CDN[5] servers like the hierarchical topology of the IP routers on the internet we believe that a hierarchical overlay network topology is required for the CDN servers to perform content routing and content delivery done efficiently.

Step 1: Try to satisfy the users request using local CDN server.

Step 2: If step 1 fails try to satisfy the user request using a CDN server inside the cluster including the local CDN server

Step 3: If step 2 fails try to satisfy the user request using a CDN server inside a nearby cluster

Step 4: If step 3 fails try to satisfy the user request using the origin server

3. Caching and content routing service capacity

Multicast over the internet was originally proposed at the network layer referred to as multicast. However after a decade of research there are still many hurdles in the deployment of IP multicast such as the lack of higher layer functionalities and scalable inter domain multicast routing protocols.

Data management – A document may be portioned into various parts permitting concurrent downloading from multiple peers[1] , the granularity and placement of these is critical.

Peer selection – The mechanism whereby a peer is selected as a server may take into account load balancing bandwidth availability and differentiate among peers who contribute more to the community.

Admission and scheduling policy – Limiting the number of concurrent downloaders and/or scheduling to provide differentiation priority among them.

Traffic – The request processes for documents along with the dynamic of how peers stay online and/or delete documents.

4. Methodology

- (a) Service capacity – since BT uses multiport downloads service capacity must be carefully defined. We estimate the service capacity as

Effective number of service = Total storage space / size

- (b) The effective number of replicates available in the system including partial downloads. Since peers may exist or delete a file upon completing a download we can estimate service capacity based on the following formula

$T * \text{Volume} - (\text{Number of finished} - \text{Number of seeds}) * \text{size} / \text{size}$

- (c) Throughput and delay of each peer

We estimate the average instantaneous throughput seen by each peer as follows

Average throughput per peer = throughput / number of downloads

(c) Table comparisons

Existing:

Capacity	CD delay	P2P delay
1.2626	0.2619	0.2676
0.2159	0.2261	4.7244
0.1960	0.1960	0
0.1757	0.1968	12.0091
0.1480	0.0.1472	0.5405
0.1360	1.1371	0.0968

Proposed:

Capacity	CD delay	P2P delay
0.13481	0.13472	0.0668
0.11180	0.11675	4.4275
0.09929	0.10119	1.9136
0.066592	0.06587	0.0758
0.03264	0.03221	1.3174
0.01649	0.01575	4.4876

We describe a set of network benchmark for measuring bandwidth , latency , software overhead which we have implemented over a wide variety of networks. We provide data from these benchmarks for both small and large message performance on many of these

benchmark for both small and large message performance on many of the super computer networks in use today and compare the performance of MPI to that of lower level network API.

Using our results, we examine various application speedups that can be achieved via network related optimizations such as overloading computations with communications, pipelining messages and the use of message packing. We provide a historical portrait of the trends in small message performance over the past 10 years.

5. Resource reservations

The site manager responsible for a resource makes a donation to a resource broker of its choosing for an agreed upon number of tokens. When a user wants to 'buy' this resource it contact the responsible resource broker and pays the required number of token and when the tokens 'validity' is checked get a ticket for that resource in return. The ticket is presented to the component manager responsible for the resource which verifies its authenticity. The component manager then gives the user access credentials for that resource. The network hierarchy has 3 main hierarchical components i.e. user equipment, Radio access network and core network.

6. Mobility management in 4G networks using IPV6

High speed – 4G[9,10] systems should offer a peak speed of more than 10 Mbits per second in stationary mode with an average of 20 Mbits per second when traveling.

High network capacity – It should be at least 10 times that of 3G[9,10] systems. This will quicken the time load time of 10 MB file to 1 second on 4G from 200 seconds on 3G enabling high speed video to stream to phone and create a virtual reality experience on high resolution handset screens fast/seamless handover across multiple networks. 4G wireless networks[8,11] should support global roaming across multiple wireless and mobile networks.

Next generation multimedia[4] support – The underlying network for 4G must be able to support fast speed and large volume of data transmission at a lower cost than today.

128 bit address space provides a sufficiently large number of address. High quality support for real time audio transmission short/ bursty connections of web applications, peer to peer[1] applications etc., Faster packet delivery, decreased cost processing no header checksum at each relay fragmentation only at the end points. Smooth hand off when the mobile host travel from one

subnet to another. Causing a change in its care of address.

7. Range of wireless services and address allocations

Broadcasting services – It include short waves like AM and FM radio as well as terrestrial television.

Mobile communication of voice and data – including maritime and aeronautical mobile for communications between ship , airplanes and land mobile for communications between a fixed base station and moving size such as a taxi fleet and paging services and mobile communications. Either between mobile user and a fixed network or between mobile user such as mobile telephone services. Fixed services either point to point or point to multi point services. Satellite used for broadcasting telecommunications and internet particularly over long distances. Other user including military radio astronomy metrological or scientific users.

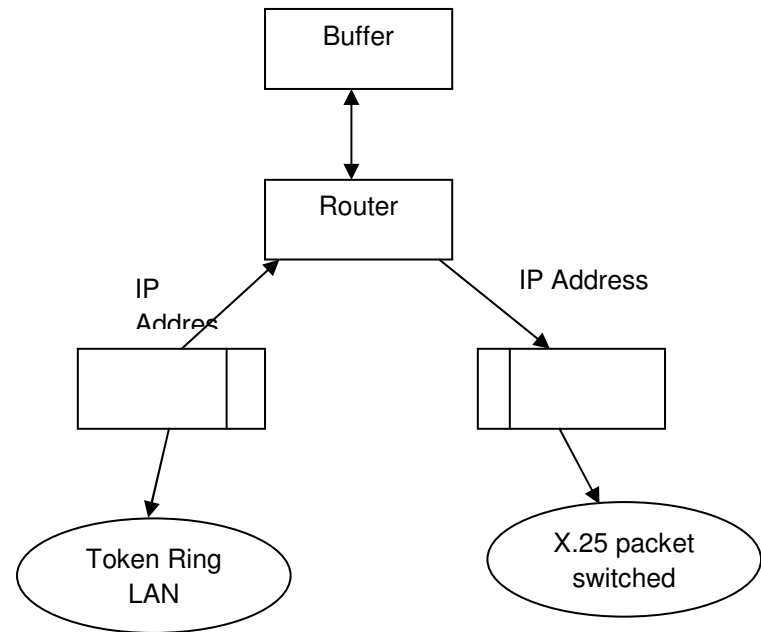
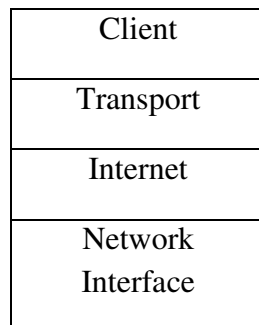


Fig 1.1

This another case that seems to duplicate lower layer features. Some errors do escape lower layer error detection. This statement may sound unusual because it implies that even if data link error detection provides reliable transmission along each link there is still no guarantee of error free transmission between the source and destination. Suppose it receives the IP packet intact but an error that affect the packet current occur during reformatting of the frame containing the packet.

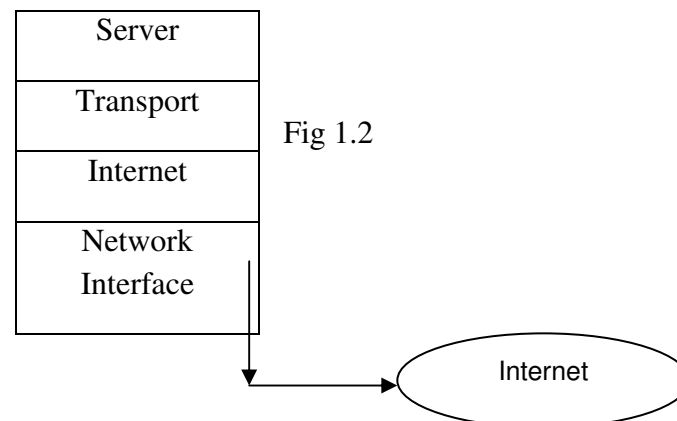


Fig 1.2

A client or server application interact directly with a transport layer protocol to establish common and to send or receive information. The transport layer protocol then uses lower layer protocol to send or receive individual messages , then a computer needs a complete stack of protocols to run either a client or a server.

Allocation space	Prefix	Fraction of address space
Reserved	0000 0000	1 / 256
Un assigned	0000 0001	1 / 256
Reserved for NSAP llocation	0000 001	1 / 128
Reserved for IPX allocation	000 010	1 / 128
Un assigned	0000 011	1 / 128
Un assigned	0000 1	1 / 32
Un assigned	0001	1 / 16
Un assigned	001	1 / 8
Provider based unicast	010	1 / 8

address		
Un assigned	011	1 / 8
Reserved for geographic based unicast address	100	1 / 8
Un assigned	101	1 / 18
Un assigned	110	1 / 8
Un assigned	1110	1 / 16
Un assigned	1111 0	1 / 32
Un assigned	1111 10	1 / 64
Un assigned	1111 110	1 / 128
Un assigned	1111 1110 0	1 / 512
Link local use address	1111 1110 10	1 / 1024
Size local use address	1111 1110 11	1 / 1024
Multicast address	1111 1111	1 / 256

8. Strategic instruments to limit competition

A call from a mobile operator to a fixed line network involves the mobile operator collecting all the revenue for the call but the fixed line network providing the termination of the call. As a result, the fixed line network usually changes for the termination of mobile to fixed calls. Typically the termination of mobile to fixed call is treated in the same way as the termination of other calls to fixed line networks being regulated on a cost basis.

A call from a fixed line network to a mobile operator often leads to an access price being set by the mobile operator. In the majority of market the termination of fixed to mobile calls has been priced many times higher than the termination of mobile to fixed calls, some times resulting in the regulation of fixed to mobile termination process. An exception is the US where this access price is the same as the mobile to fixed access price due to the symmetry requirements under the 1996 telecom act.[2,3] Other exceptions are Canada, Hong Kong and Singapore where there are no fixed to mobile termination changes. Instead like the US mobile operators[4] recover the cost of terminating fixed to mobile calls from the mobile callers directly.

A call from a mobile operator to another mobile operator may involve mobile operator charging each other

for termination. Such access prices are often negotiated privately between mobile operators.

Wayne Thomson – SAS has been one of the first vendors to recognize the need for operationalizing analytics i.e. moving beyond just knowledge discovery. Open standard have been very critical to deploy analytical models in a heterogeneous environment.

Mike Hockins – We will be announcing our data rush based parallel data mining solution for the summer. Shortly thereafter we expect to announce support for reading and writing PMML both at design time and run time.

Robert Grossmann – I think cloud computing[10] is the disruptive technology that will have the greatest impact on data mining and predictive model in the near term.

9. Metrological and scientific issues in internet

Processor in the cluster may not be identical. The communication network may have a regular but heterogeneous structure. The cluster may be a multi-user computer system. This is in particular makes the performance character of the processor dynamic and non identical. Mobile telecom system with different type of processor from embedded into mobile phones to central computers processing calls.

Embedded control multiprocessor system ex. Car, airplanes etc., Given a set of processors the speed of each of which is characterized by a positive constant.

Partition is the mathematical object into sub objects of the same type. There is one to one mapping between the position and the processors. The size of each partition is proportional to the speed of the processor owing the partition. Some additional restrictions to the relationship between the partitions are satisfied. The partitioning minimizes some functional, which is used to measure each partitioning. The whole computation is partitioned into a number of equal chunks. Each chunk is performed by separate process. The number of process run by each processor is proportional to the relative speed of the processor.

The performance of heterogeneous model network of computers quantifying the ability of the network to perform computations and communications. Performance model of applications quantifying the computations and communications to be performed by the application. Performance model of a set of heterogeneous processor which is used to estimate the execution time of computations. Performance model of communication network which is necessary to predict the execution time of communication operations. In the operation of external load and

paging all processors are supposed to demonstrate the same speed. Each processor is characterized by the two parameters i.e. Current CPU utilizations and the size of current available memory.

The serial code is provided by the application programmer. It is supposed that the code is representative for the computations performed during the execution of the application. The code is performed at runtime in the points on the application specified by the application programmer. It is supposed that the code is representative for the computations performed during the execution of the application. The code is performed at runtime in the points of the application specified by the application programmer. Thus the performed model of the processor provides current estimation of their speed demonstrated on the code representative for the particular application. The number of process executing the algorithm which normally of the parameter the model. The total volume of the computations to be performed by each processor during the execution of the algorithm. The total volume of data transferred between each pair of the process during the execution of the algorithm.

Currently owned and used space from the data user ISP's. New space obtained by directly from internet

registries. Customer space private autonomous system currently owned and used data center. A new AS obtained directly from customer internet registries. One to one, One to many, Many to one and Many to many relationships are possible. 25% of the new space must be utilized immediately and 50% of the space is used within one year. To get more space and the added space must be utilized as 80%. The private IP address conflict with other customer networks. When the network communication is initiated and where the application that need to be run and that can be work across NAT boundaries.[10,11]

Conclusion

In this paper we presented an alternative theory of the internet evolution and developed a new topology generator based on wealth evolution and random walks. Our method reconfigures the VNT gradually by dividing it into multiple stages. This would reduce dramatically the sites of matrices whose pseudo inverse matrices needed to be calculated at each stage. The eventual goal is the practical design of each scheme to achieve the performance outlined in this paper.

Abbreviations

ISP – Internet service Protocol

RFC – Requested flow control

CDN – Content delivery networks

MPI – Multi Program Interface

API – Application program interface

SAS – Source address service

PMMI – Phase modulation mobile interface

NAT – Network address technology

VNT – Virtual network terminal

References

1. Analysis of peer to peer internet telephony protocols by Salman A Basket
2. Cyber security threats and counter measures by Buheita Fujiwara
3. Network security policy by Frederic M Avolio July 2007
4. Multimedia content delivery networks by Jian Ni and H K Tsang
5. Evaluation of current high performance networks by Paul Hargrove
6. Enhancing the security of Wi-Fi networks by Paramvir Bahl
7. Wireless sensor networks by F L Lewis
8. Flexible resource allocation in 3G networks by Mohamed Al-Fares
9. Wireless communications by Joshua S Gans
10. Open standard and cloud computing by Michael Zeller Zementics
11. High performance grid computing by Jack Dongarra

Framework for Customized-SOA Projects

Sheikh Muhammad Saqib¹, Muhammad Zubair Asghar¹, Shakeel Ahmad¹, Bashir Ahmad¹ and Muhammad Ahmad Jan¹

¹Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan

Abstract—Presentation, business logic and information access are becoming modern trends for developing SOA applications. Building blocks of SOA include logical grouping of components to accomplish business functionality. Although Enterprise SOA application has no. of development methodologies, but CSOA (Customized SOA) application development is still at beginner level. Customized SOA application is just like customized software for an organization with no entry of physical material except logical material. Due to common characteristics of RUP and SOA, authors proposed a Business Modeling based model for CSOA. Proposed model for CSOA is a development framework which uses the discipline & practices of SOMA and RUP for the proper management of utilizing the basic characteristics of SOA such as reusing & changing existing service with less effort, adding new service at any time, developing new services with future plan and etc. Author suggests that proposed model will be the best development framework for COSA with respect to rest of models.

Keywords- SOA, CSOA, SOMA, RUP and Service.

I. INTRODUCTION

An authority that combines business services to business practices all together leads to 'SOA is a software application' [1]. Analysis, modeling, design techniques, and activities of SOMA Service-oriented modeling architecture is the foundations of a SOA. SOAM is used to identify, specify and realize the services [2]. RUP Rational Unified Process is a development framework for effective project management. Managing requirements, component-based architecture, continuously verify quality and control changes and visually model are the different best practices for RUP model [3]. Major purpose of author [4] work is to improve SOA plug-in for RUP. Service oriented modeling and architecture work is similar to four phases of RUP i.e. inception, elaboration, construction and transition. Because activities of RUP last phases is working same service identification, specification and realization. During the inception phase, emphasis is given to determine the scope of project not in term of services. In normal development of RUP project, inception is considered as core phase to understand the whole concept but in SOA methodology inception can be namely considerable piece of the service realization activity [4]. Enterprise SOA application has no of development methodologies such as SOMA, but CSOA (Customized SOA) application development is still at beginner level. Customized SOA application is just like customized software for an organization with no or minimum entry of physical material except logical material. It provides only services to an organization according to their needs. CSOA needs a way for development method as it is application for small organization, whereas there exist many methodologies

for developing Enterprise level application such as SOMA and RUP. SOMA is a method for designing and implementing enterprise SOA based application while RUP is used for huge project [1] [5]. Each methodology has its own confines according to their nature. SOMA or RUP can be used for developing CSOA applications with the fact that effort, cost and time may not be pleasing for CSOA due to its small volume. Considering CSOA constraints, a method is needed for its development to overcome limitations associated with RUP or SOMA. Common characteristics of SOA and RUP such as size of development team, level of documentation, development time and type of orientation [1] enabled us to propose a development approach for developing CSOA with coordination of SOMA and RUP, because three RUP phases contains all activities of specification, identification and realization. Work of inception is distributed among all these activities [4].

II. SOMA & RUP

In valid globe, Service implies a piece of independent deployment and versioning, communicate with other components and encapsulated [6]. SOA got importance due to different features such as loosely coupled components, black boxes services and SOA service should be self defined. SOA architecture includes: SOA registry, SOA workflow, Service broker, and SOA supervisor [7].

Service Oriented Modelling and Architecture consider as a best attitude for SOA. It is a method to plan and implement a SOA. It contains different phases for development of SOA. Each consists of many activities comprising compositions of tasks, which can be executed by roles. SOMA method describes all activities, roles and work products needed to design and implement a SOA [2]. SOMA phases are: Service Identification: identifies candidate service. Service Specification: specify the services sufficiently detailed to develop them. Service Realization: how services will be implemented and developed. Service Implementation: Service components are written, wrappers and services are assembled. Service Deployment: handles everything related to the deployment of the SOA to the customer [2].

Within development organization, a software engineering process model RUP, is a disciplined approach which assigns different responsibilities and tasks. Under the predictable schedule and budget RUP creates valuable software according to user's requirements [3].

III. PROPOSED MODEL FOR CSOA

Like Service Oriented Modeling and Architecture, a model for CSOA will also require identification, specification, realization and deployment. These four sections will use the

different practices of RUP and SOMA, then customize based SOA applications can easily incorporate all characteristics of enterprise based SOA projects.

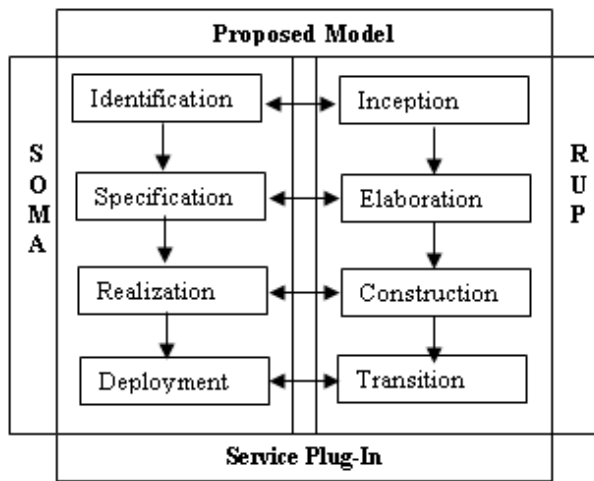


Figure 1: Flow of proposed model.

In above Fig-1, it is clear that proposed model will also have four phases to construct CSOA applications.

A. Identification of Customized Services

RUP business modeling discipline is useful for determining the view of business structure, processes of the organization, identifying scope of the project, but it is not sufficient for the enterprise, because it does not depict cross-system issues adequately [8]. Since CSOA is a customized level application, then Business Modeling can be easily applied on CSOA. Analysis of this activity will be coarse-grained.

SOMA Domain Decomposition is a top-down method that analyzes business domains and business use cases to identify services and how the business use cases are executed in detail and this analysis will be fine-grained i.e. determining a service to a specific function [2], but here we will use story cards instead of use cases. Story cards can be created by means of coordination of users (already identified) and system analyst. With the help of Story Card Headers and business glossary, each story cards can be easily created and setting the priorities of different services.

Its major purpose will explore the business objective and domain decomposition by technically arranging the following practices.

- Description of customer business state.
- Description of different boundaries for modeling.
- Separation of services with respect to category i.e. general, specific or more specific.
- Scope of services for future.
- Reused and reusable services.
- Availability of candidate Services
- Technical terms used in business domain.
- Mention Users for writing and evaluating story cards.

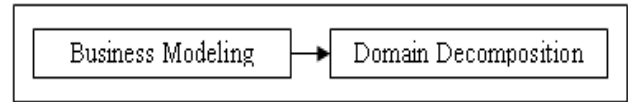


Figure 2: Work of 1st phase.

B. Specification of Customized Services

Gathered requirements about the product must be understood by the end of this phase. If there is no chance for using pre-built framework going to be used for construction then elaboration phase is considered complex. First phase results will be very useful to elaborate each service specification [9].

Objective of this phase is to elaborate different specification about services. Goal of this phase is to specify those services which should be implemented during this iteration. Following technical practices will be arranged in this phase.

- Specification of functional components of a service.
- Exploring different components attribute.
- Applying coding standard and metaphor on above components.
- Class diagrams from above components.

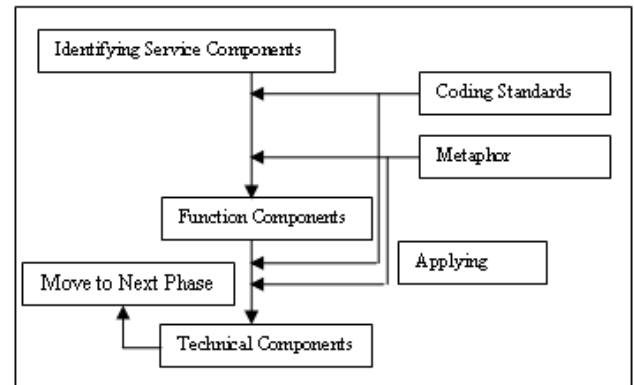


Figure 3: Flow of 2nd phase.

C. Realization of Customized Services

This phase will control the operation, management of resources & optimize the schedules, costs, and quality. Its major theme is analogous to construction phase of RUP and service realization of SOMA.

Its major theme is analogous to construction phase of RUP and service realization of SOMA. Its major views are: To determine how services will be implemented, How service will be constructed, User manuals creation. Current released report. Following two practices will be done when service will already exist in service library.

- Mapping of analyzed services with existing services.
- Changing existing services if some needs are required.

- Unit testing (Verification of functionality of a specific section of code).
- Plug-in service to system.

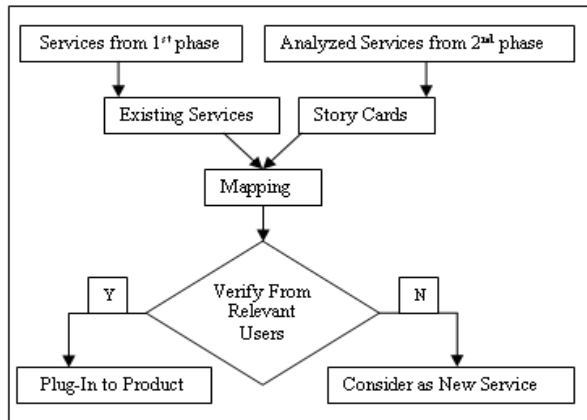


Figure 4: Flow of 3rd phase for existing services.

If service does not exist in library then following practices will be done.

- Technical exploration of a service.
- Service communication interface (visible or invisible interface to end user).
- Unit testing.
- Plug-in service to system.

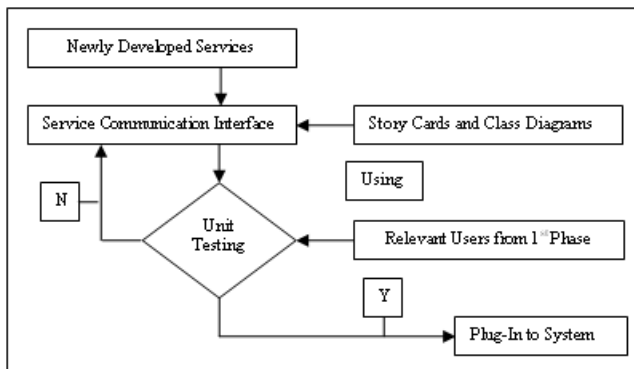


Figure 5: Flow of 3rd phase for newly developed services.

D. Deployment of Customized Services

Plug-in the services in productive environment and taking the user acceptance tests will be the major objective of this phase. This includes:

Major goal of this phase is

- To plug-in the services in productive environment.
- Taking the user acceptance tests.
- Beta Testing (To validate the new whole system against user expectations).
- Conversion of operational databases.
- Users Training.

- Store reusable services and its manual in library.

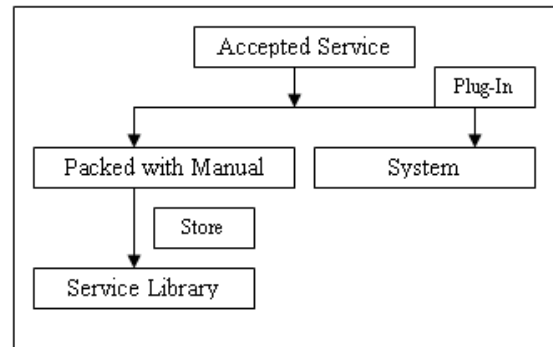


Figure-6: Flow of 4th phase for newly developed services.

IV. CONCLUSION AND RESULTS

Customized SOA application can be developed with proposed model more efficiently as compared to rest of models for customized applications. Some models use sequential approach, some use iterative approach and some uses reusability concepts. Since CSOA is a loosely coupled application and can't be created with respect to traditional models. Proposed model is composed of practices from RUP (Business Oriented) and SOMA (Service Oriented), so it can be named as Business Oriented Service Model.

Proposed model can have following characteristics:

Worth of service

- Ease of requirements understanding and writing.
- Ease to estimate development efforts.
- Easy to change and adding new Service.
- Evaluation on spot.
- Well defined executable architectural prototype.
- Integrated existing services will be more accurate.

REFERENCES

- [1] Guillermo A. Callahan; Suitability of Extreme Programming and RUP Software Development Methodologies for SOA Applications, 2006.
- [2] Bachelor Thesis on Model-driven Software Migration into a Service-oriented Architecture: Dr. Hans-Jürgen, 2009.
- [3] Shahid Hussain: Role Communication Model in RUP: Using Pair Programming Concepts, 2008.
- [4] Hans Admiraal: SOA gains importance to the RUP, (2008).
- [5] How to Adopt RUP in Your Project, URL:
www.pearsonhighered.com/samplechapter/0321202945.pdf 2003
- [6] "What is a Service in an SOA?", 2007 URL: <http://soa-eda.blogspot.com/2007/01/what-is-service-in-soa.html>
- [7] SOA : Service Oriented Architecture:
url:<http://www.dotnetspark.com/kb/271-service-oriented-architecture-soa-terview.aspx>, 2009
- [8] The Enterprise Business Modeling Discipline: Scaling Agile Software Development, URL:
<http://www.enterpriseunifiedprocess.com/essays/enterpriseBusinessModeling.html>
- [9] "The Furious Five", URL: www.jmu.edu/orgs/achesa/models.doc

About Author

SHEIKH MUHAMMAD SAQIB is with Gomal University, D.I.Khan, Pakistan. (E-mail: saqibsheikh4@hotmail.com). His research interest is in RUP, SOA and Cloud Computing. Some of his papers were presented in some conference such as 9th National Research Conference at SZABIST – Islamabad PAKISTAN, 2009, ISCS (International Symposium on Computer Science) at Namal College Mianwali PAKISTAN an Associate College of The University of Bradford, UK, 2011 and Ten (10) are published in international journals.

A New Hybrid Neural Model for Real-Time Prediction Applications

Hazem M. El-Bakry,

Faculty of Computer Science & Information Systems,
Mansoura University,
Mansoura,
EGYPT
E-mail: helbakry20@yahoo.com

Wael . A. Awad

Math. & Comp. Dept., Faculty of Science,
Port Said University,
Port Said,
EGYPT

Abstract- In this paper, a new hybrid neural model for prediction applications is presented. Fast Feedforwarded Neural Networks (FFNNs) is integrated with modified recurrent neural networks for powerful estimation. The proposed new model is applied for prediction of power consumption. First, Modified Kohonen's Neural Networks (MKNNs) are used to facilitate the prediction process because they have the ability for clustering the input space into a number of classes. Therefore it is used for data classification to identify the categories which are essential for the prediction process. The unsupervised process performs the role of front-end data compression. For each category, the supervised learning algorithm LVQ is used for training FFNNs. The operation of FFNNs relies on performing cross correlation in the frequency domain between the input data and the weights of neural networks. Simulation results have shown that the presented integrated neural model is very powerful prediction.

Keywords- FFNNs, MKNNs, LVQ, Cross Correlation, Frequency Domain and Prediction

I. INTRODUCTION

A prediction is a statement about the way things will happen in the future, often but not always based on experience or knowledge. While there is much overlap between *prediction* and *forecast*, a *prediction* may be a statement that some outcome is expected, while a *forecast* may cover a range of possible outcomes. Although guaranteed information about the information is in many cases impossible, prediction is necessary to allow plans to be made about possible developments; Howard H. Stevenson writes that prediction in business "... is at least two things: Important and hard [53]. Prediction is closely related to uncertainty. Reference class forecasting was developed to eliminate or reduce uncertainty in prediction [54].

Outside the rigorous context of science, prediction is often confused with informed guess or opinion. A prediction of this kind might be (inductively) valid if the predictor is a knowledgeable person in the field and is employing sound reasoning and accurate data. Large corporations invest heavily in this kind of activity to

help focus attention on possible events, risks and business opportunities, using futurists. Such work brings together all available past and current data, as a basis to develop reasonable expectations about the future. Predictions have often been made, from antiquity until the present, by using paranormal or supernatural means such as prophecy or by observing omens. Methods including water divining, astrology, numerology, fortune telling, interpretation of dreams, and many other forms of divination, have been used for millennia to attempt to predict the future. These means of prediction have not been substantiated by controlled experiments, and are disputed by most, including scientists and skeptics.

In politics it is common to attempt to predict the outcome of elections via political forecasting techniques (or assess the popularity of politicians) through the use of opinion polls. Prediction games have been used by many corporations and governments to learn about the most likely outcome of future events. Predictions have often been made, from antiquity until the present, by using paranormal or supernatural means such as prophecy or by observing omens. Methods including water divining, astrology, numerology, fortune telling, interpretation of dreams, and many other forms of divination, have been used for millennia to attempt to predict the future. These means of prediction have not been substantiated by controlled experiments, and are disputed by most, including scientists and skeptics.

In science a prediction is a rigorous, often quantitative, statement, forecasting what will happen under specific conditions; for example, if an apple falls from a tree it will be attracted towards the center of the earth by gravity with a specified and constant acceleration. The scientific method is built on testing assertions that are logical consequences of scientific theories. This is done through repeatable experiments or observational studies. A scientific theory whose assertions are contradicted by observations and evidence will be rejected. Notions that make no *testable* predictions are usually considered not to be part of science (protoscience or nescience) until testable predictions can be made. New

theories that generate many new predictions can more easily be supported or falsified (see predictive power). In some cases the probability of an outcome, rather than a specific outcome, can be predicted, for example in much of quantum physics. Mathematical equations and models, and computer models, are frequently used to describe the past and future behaviour of something. In microprocessors, branch prediction permits avoidance of pipeline emptying at branch instructions. In engineering, possible failure modes are predicted and avoided by correcting the mechanism causing the failure. Accurate prediction and forecasting are very difficult in some areas, such as software reliability, natural disasters, pandemics, demography, population dynamics and meteorology.

Established science makes useful predictions which are considered to be extremely reliable and accurate; for example, eclipses are routinely predicted. New theories make predictions which allow them to be falsified if the predictions are not borne out. For example in the early twentieth century the scientific consensus was that there was an absolute frame of reference, given the name *luminiferous ether*. The famous Michelson-Morley experiment ruled this out, falsifying the idea of an absolute frame and leaving the very counter-intuitive special theory of relativity as the only possibility.

Albert Einstein's theory of general relativity could not easily be tested as it did not produce any effects observable on a terrestrial scale. However, the theory predicted that large masses such as stars would bend light, in contradiction to accepted theory; this was observed in a 1919 eclipse. Mathematical models of stock market behaviour are also unreliable in predicting future behaviour. Consequently, stock investors may anticipate or predict a stock market boom, or fail to anticipate or predict a stock market crash. Some correlation has been seen between actual stock market movements and prediction data from large groups in surveys and prediction games.

An actuary uses actuarial science to assess and predict future business risk, such that the risk(s) can be mitigated. For example, in insurance an actuary would use a life table to predict (truly, estimate or compute) life expectancy. In literature, vision and prophecy are literary devices used to present a possible timeline of future events. They can be distinguished by vision referring to what an individual sees happen. The New Testament book of Revelation (Bible) thus uses vision as a literary device in this regard. It is also prophecy or prophetic literature when it is related by an individual in a sermon or other public forum. Fiction (especially fantasy, forecasting and science fiction) often features instances of prediction achieved by unconventional means.

- In fantasy literature, predictions are often obtained through magic or prophecy, sometimes referring back to old traditions. For example, in J. R. R. Tolkien's *The Lord of the Rings*, many of the characters possess an awareness of events

extending into the future, sometimes as prophecies, sometimes as more-or-less vague 'feelings'. The character Galadriel, in addition, employs a water "mirror" to show images, sometimes of possible future events.

- In some of Philip K. Dick's stories, mutant humans called *precogs* can foresee the future (ranging from days to years). In the story called *The Golden Man*, an exceptional mutant can predict the future to an indefinite range (presumably up to his death), and thus becomes completely non-human, an animal that follows the predicted paths automatically. Precogs also play an essential role in another of Dick's stories, *The Minority Report*, which was turned into a film by Steven Spielberg in 2002.
- In the *Foundation* series by Isaac Asimov, a mathematician finds out that historical events (up to some detail) can be theoretically modelled using equations, and then spends years trying to put the theory in practice. The new science of psychohistory founded upon his success can simulate history and extrapolate the present into the future.
- In Frank Herbert's sequels to *Dune*, his characters are dealing with the repercussions of being able to see the possible futures and select amongst them. Herbert sees this as a trap of stagnation, and his characters follow a Golden Path out of the trap.
- In Ursula K. Le Guin's *The Left Hand of Darkness*, the humanoid inhabitants of planet Gethen have mastered the art of prophecy and routinely produce data on past, present or future events on request. In this story, this was a minor plot device.

FFNNs for detecting a certain code in one dimensional serial stream of sequential data were described in [10,27]. Compared with traditional feedforward neural networks (TFNNs), FFNNs based on cross correlation between the tested data and the input weights of neural networks in the frequency domain showed a significant reduction in the number of computation steps required for certain data detection [5-35]. Here, we make use of the theory of FFNNs implemented in the frequency domain to increase the speed of neural networks for prediction of power consumption. The idea of moving the testing process from the time domain to the frequency domain is applied to time delay neural networks. Theoretical and practical results show that the proposed FFNNs are faster than TFNNs.

This paper is organized as follows. Prediction of power consumption by using ANNs is discussed in section II. The Theory of FFNNs is described in section III. The proposed approach for prediction is presented in section IV. Finally conclusions are given.

II. PREDICTION OF POWER CONSUMPTION BY USING ANNS

Artificial neural network (ANNs) is a mathematical model, which can be set one or more layered and occurred from many artificial neural cells. The wide usage of the ANN may be due to the three basic properties:

1. The ability of the ANNs as a parallel processing of the problems, for which if any of the neurons violate the constraints would not affect the overall output of the problem.
2. The ability of the ANNs to extrapolate from historical data to generate forecasts.
3. The successful application of the ANN to solve non-linear problems. The history and theory of the ANN have been described in a large number of published literatures and will not be covered in this paper except for a very brief overview of how neural networks operate.

The ANN computation can be divided into two phases: learning phase and testing phase. The learning phase forms an iterative updating of the synoptic weights based upon the error back propagation algorithm. Back propagation algorithm is generalized of least mean square learning rule, which is an approximation of steepest descent technique. To find the best approximation, multi-layer feed forward neural network architecture with back propagation learning rule is used. A schematic diagram of typical multi-layer feed-forward traditional neural network architecture. The number of neurons in the hidden layer is varied to give the network enough power to solve the forecasting problem. Each neuron computes a weighted sum of the individual inputs (I_1, I_2, \dots, I_j) it receives and adding it with a bias (b) to form the net input (x). The bias is included in the neurons to allow the activation function to be offset from zero.

ANNs is a computing system that imitates intelligent behavior and processes information by its dynamic state response to external inputs. The neural network has two processes; learning and classification. There are two types of learning; supervised and unsupervised [1-5]. In the supervised learning process the network is supplied with a pair of patterns, an input pattern and a corresponding target output pattern. In the classification process, the user provides a pattern of system description parameters to the neural network and it returns an estimate of the output pattern. The unsupervised learning's used for clustering the input space into affinity number of classes represented by the neural network weights vectors. A principal challenge in neural computing is adjusting the set of internal parameters known as weights in order to encode an underlying relation assumed to exist amongst various data sets.

The power consumption (Gas, Electricity, Water) demand is influenced by many factors, such as weather, economic and social activities and different power consumption components (residential, industrial, commercial etc.). By analysis of only historical power consumption data, it is difficult to obtain accurate power consumption demand for prediction. The relation between power consumption demand and the independent variables is complex and it is not always possible to fit the

power consumption curve using statistical models [6]. The daily power consumption patterns in the same geographical area have been repeated for the same day type in the same season so, the ANNs approach was proposed for power consumption prediction [7-13].

Sophisticated algorithms more or less rely on presumptions on influences like temperature, humidity, wind speed, cloudiness, and rain. Although most of these information might be taken and updated (automatically) from the statistics applied to recorded data, the underlying power consumption models inherited in power consumption prediction algorithms make traditional approaches inflexible against major changes in the consumer behavior. In order to overcome these problems, AI techniques have been applied to power consumption prediction. The ANNs approach has several key features that make it highly suitable for power consumption prediction [12]. For example,

- It does not require any presumed functional relationship between electric power consumption and other variables such as weather conditions.
- It provides a nonlinear mapping between weather variables and previous power consumption patterns, and electric power consumption with out the need for predetermined model.
- It is usually fault tolerant and robust.

ANNs proved to be capable of finding internal representations of interdependencies within raw data not explicitly given or even known by human experts. This typical characteristic together with the simplicity of building and training ANNs and their very short response time encouraged various groups of researchers to apply ANNs to the task of power consumption prediction. Most of the papers [6-13] present feasibility studies carried out at universities or research institutes often in cooperation with utilities. They mainly addresses peak power consumption prediction, total power consumption prediction, and hourly power consumption prediction with lead time from 1 to 48 hours. Typically ANNs map input data to predict the value of power consumption. Therefore, FFNNs are used. It has been proved that these networks are very efficient in many different applications [1-37]. The self-organizing feature map is used by [8] which clusters the prediction data to clusters with similar power consumption profiles. There is no paper describing an ANNs concept coping with the task of e.g. hourly power consumption prediction for all days of the year. In this paper, a generalized approach based on ANNs for prediction process is presented. An Information system that Implement the generalized approach is developed. MKNNs is used for data classification to identify the day classes/types which are essential for prediction processes.

Due to the dynamic nature of hourly power consumptions and differences in power consumption characteristics from region to region, it is required to analysis power consumption data for each region separately and to design suitable neural networks for prediction process for that region. Power consumption

analysis is performed to identify the day type/classes of that region. In this work, it is required to collect the effective hourly data of the power system and save them in a dynamic database. Once the database is developed, their information can be update, delete, display, and insert. Also, from database we can identify the day types/classes of daily power consumption patterns using MKNNs.

III. THEORY OF FFNNs FOR PREDICTION

Computing the resulted output; for a certain pattern of information; in the incoming serial data, is a prediction problem. First neural networks are trained to predict the estimated variable and this is done in time domain. In pattern detection phase, each position in the incoming matrix is processed to predict the estimated variable by using neural networks. At each position in the input one dimensional matrix, each sub-matrix is multiplied by a window of weights, which has the same size as the sub-matrix. The outputs of neurons in the hidden layer are multiplied by the weights of the output layer. Thus, we may conclude that the whole problem is a cross correlation between the incoming serial data and the weights of neurons in the hidden layer. The convolution theorem in mathematical analysis says that a convolution of f with h is identical to the result of the following steps: let F and H be the results of the Fourier Transformation of f and h in the frequency domain. Multiply F and H^* in the frequency domain point by point and then transform this product into the spatial domain via the inverse Fourier Transform. As a result, these cross correlations can be represented by a product in the frequency domain. Thus, by using cross correlation in the frequency domain, speed up in an order of magnitude can be achieved during the detection process [5-35]. Assume that the size of the attack code is $1 \times n$. In attack detection phase, a sub matrix I of size $1 \times n$ (sliding window) is extracted from the tested matrix, which has a size of $1 \times N$. Such sub matrix, which may be an attack code, is fed to the neural network. Let W_i be the matrix of weights between the input sub-matrix and the hidden layer. This vector has a size of $1 \times n$ and can be represented as $1 \times n$ matrix. The output of hidden neurons $h(i)$ can be calculated as follows [5]:

$$h_i = g \left(\sum_{k=1}^n W_i(k) I(k) + b_i \right) \quad (1)$$

where g is the activation function and $b(i)$ is the bias of each hidden neuron (i). Equation 1 represents the output of each hidden neuron for a particular sub-matrix I . It can be obtained to the whole input matrix Z as follows [5]:

$$h_i(u) = g \left(\sum_{k=-n/2}^{n/2} W_i(k) Z(u+k) + b_i \right) \quad (2)$$

Eq.1 represents a cross correlation operation. Given any two functions f and d , their cross correlation can be obtained by [4]:

$$d(x) \otimes f(x) = \left(\sum_{n=-\infty}^{\infty} f(x+n) d(n) \right) \quad (3)$$

Therefore, Eq. 2 may be written as follows [5]:

$$h_i = g(W_i \otimes Z + b_i) \quad (4)$$

where h_i is the output of the hidden neuron (i) and $h_i(u)$ is the activity of the hidden unit (i) when the sliding window is located at position (u) and $(u) \in [N-n+1]$.

Now, the above cross correlation can be expressed in terms of one dimensional Fast Fourier Transform as follows [5]:

$$W_i \otimes Z = F^{-1} (F(Z) \bullet F^*(W_i)) \quad (5)$$

Hence, by evaluating this cross correlation, a speed up ratio can be obtained comparable to traditional neural networks. Also, the final output of the neural network can be evaluated as follows:

$$O(u) = g \left(\sum_{i=1}^q W_o(i) h_i(u) + b_o \right) \quad (6)$$

where q is the number of neurons in the hidden layer. $O(u)$ is the output of the neural network when the sliding window located at the position (u) in the input matrix Z . W_o is the weight matrix between hidden and output layer.

The complexity of cross correlation in the frequency domain can be analyzed as follows:

1- For a tested matrix of $1 \times N$ elements, the 1D-FFT requires a number equal to $N \log_2 N$ of complex computation steps [13]. Also, the same number of complex computation steps is required for computing the 1D-FFT of the weight matrix at each neuron in the hidden layer.

2- At each neuron in the hidden layer, the inverse 1D-FFT is computed. Therefore, q backward and $(1+q)$ forward transforms have to be computed. Therefore, for a given matrix under test, the total number of operations required to compute the 1D-FFT is $(2q+1)N \log_2 N$.

3- The number of computation steps required by FFNNs is complex and must be converted into a real version. It is known that, the one dimensional Fast Fourier Transform requires $(N/2) \log_2 N$ complex multiplications and $N \log_2 N$ complex additions [3]. Every complex multiplication is realized by six real floating point operations and every complex addition is implemented by two real floating point operations. Therefore, the total number of computation steps required to obtain the 1D-FFT of a $1 \times N$ matrix is:

$$\rho = 6((N/2) \log_2 N) + 2(N \log_2 N) \quad (7)$$

which may be simplified to:

$$\rho = 5N \log_2 N \quad (8)$$

4- Both the input and the weight matrices should be dot multiplied in the frequency domain. Thus, a number of complex computation steps equal to qN should be considered. This means $6qN$ real operations will be added to the number of computation steps required by FFNNs.

5- In order to perform cross correlation in the frequency domain, the weight matrix must be extended to have the same size as the input matrix. So, a number of zeros $= (N-n)$ must be added to the weight matrix. This requires a total real number of computation steps $= q(N-n)$ for all neurons. Moreover, after computing the FFT for the weight matrix, the conjugate of this matrix must be obtained. As a result, a real number of computation steps $= qN$ should be added in order to obtain the conjugate of the weight matrix for all neurons. Also, a number of real computation steps equal to N is required to create butterflies complex numbers $(e^{-jk(2\pi n/N)})$, where $0 < K < L$. These $(N/2)$ complex numbers are multiplied by the elements of the input matrix or by previous complex numbers during the computation of FFT. To create a complex number requires two real floating point operations. Thus, the total number of computation steps required for FFNNs becomes:

$$\sigma = (2q+1)(5N \log_2 N) + 6qN + q(N-n) + qN + N \quad (9)$$

which can be reformulated as:

$$\sigma = (2q+1)(5N \log_2 N) + q(8N-n) + N \quad (10)$$

6- Using sliding window of size $1 \times n$ for the same matrix of $1 \times N$ pixels, $q(2n-1)(N-n+1)$ computation steps are required when using TFNNs for certain attack detection or processing (n) input data. The theoretical speed up factor η can be evaluated as follows:

$$\eta = \frac{q(2n-1)(N-n+1)}{(2q+1)(5N \log_2 N) + q(8N-n) + N} \quad (11)$$

Time delay neural networks accept serial input data with fixed size (n) . Therefore, the number of input neurons equals to (n) . Instead of treating (n) inputs, the proposed new approach is to collect all the incoming data together in a long vector (for example $100 \times n$). Then the input data is tested by time delay neural networks as a single pattern with length L ($L=100 \times n$). Such a test is performed in the frequency domain as described before.

The theoretical speed up ratio for searching short successive (n) code in a long input vector (L) using time delay neural networks is listed in tables I, II, and 3. Also, the practical speed up ratio for manipulating matrices of different sizes (L) and different sized weight matrices (n) using a 2.7 GHz processor and MATLAB is shown in table IV. An interesting point is that the memory capacity is reduced when using FFNNs. This is because the number of variables is reduced compared with TFNNs.

IV. THE PROPOSED TECHNIQUE FOR PREDICTION OF POWER CONSUMPTION

From the analysis of the prediction methods [6-11], a generalized approach for prediction using ANNs is concluded. The steps of the generalized approach for prediction process using ANNs can be described as:-

Phase 1: Data Collection

Form a historical database that contains the data of the attributes of the prediction process (input-output) that cover enough period immediately preceding the current time.

Phase 2: Data Classification

- Classify the historical database into groups according to a certain criteria (day type, season, similarity,...)
- Reject the redundant and inconsistent records from the database.
- Identify the prediction parameters (period, class, ...).

Phase 3: Training and testing

Identify the initial design of FFNNs that requires to determine the following parameters (no of input, output, hidden neurons, no of hidden layers, activation function, learning rate and momentum rate, and no of iterations).

- Form training and test sets of patterns for a class or day type from the database.
- Normalize the training and test sets.
- Train the neural network using the suitable learning algorithm.
- Test the trained FFNNs
- Calculate the testing performance measures of the FFNNs (absolute percentage error , average error and standard deviation)
- If the performance measures are not accepted then change the FFNNs design parameters and repeat phase 3.

Phase 4: Saving and updating

- Save the parameters of the trained FFNNs and its weights (Final ANNs design).
- Update the historical database with the current recorded values.

In this paper, the database is used to form the training patterns for the MKNNs. Phase 2 is performed with MKNNs. The output of phase 2 is the classes of data patterns that corresponding to day types classification. Phase 3 and 4 are presented in the second part of this work. An extra-large supervised learning system and a hierarchical system [10] are shown in Figures 1 and 2. Fig.1 illustrates a situation where in almost every cluster there are a few local regions and separate supervised learning procedures are necessary. In Fig. 2 large numbers of training set patterns are then sent forward to a supervised

TABLE I. THE THEORETICAL SPEED UP RATIO FOR PREDICTION OF POWER CONSUMPTION (n=400).

Length of serial data	Number of computation steps required for TFNNs	Number of computation steps required for FFNNs	Speed up ratio
10000	2.3014e+008	4.2926e+007	5.3613
40000	0.9493e+009	1.9614e+008	4.8397
90000	2.1478e+009	4.7344e+008	4.5365
160000	3.8257e+009	8.8219e+008	4.3366
250000	5.9830e+009	1.4275e+009	4.1912
360000	8.6195e+009	2.1134e+009	4.0786
490000	1.1735e+010	2.9430e+009	3.9876
640000	1.5331e+010	3.9192e+009	3.9119

TABLE II. THE THEORETICAL SPEED UP RATIO FOR PREDICTION OF POWER CONSUMPTION (n=625).

Length of serial data	Number of computation steps required for TFNNs	Number of computation steps required for FFNNs	Speed up ratio
10000	3.5132e+008	4.2919e+007	8.1857
40000	1.4754e+009	1.9613e+008	7.5226
90000	3.3489e+009	4.7343e+008	7.0737
160000	0.5972e+010	8.8218e+008	6.7694
250000	0.9344e+010	1.4275e+009	6.5458
360000	1.3466e+010	2.1134e+009	6.3717
490000	1.8337e+010	2.9430e+009	6.2306
640000	2.3958e+010	3.9192e+009	6.1129

TABLE III. THE THEORETICAL SPEED UP RATIO FOR PREDICTION OF POWER CONSUMPTION (n=900).

Length of serial data	Number of computation steps required for TFNNs	Number of computation steps required for FFNNs	Speed up ratio
10000	4.9115e+008	4.2911e+007	11.4467
40000	2.1103e+009	1.9612e+008	10.7600
90000	4.8088e+009	4.7343e+008	10.1575
160000	0.8587e+010	8.8217e+008	9.7336
250000	1.3444e+010	1.4275e+009	9.4178
360000	1.9381e+010	2.1134e+009	9.1705
490000	2.6397e+010	2.9430e+009	8.9693
640000	3.4493e+010	3.9192e+009	8.8009

TABLE IV. PRACTICAL SPEED UP RATIO FOR PREDICTION OF POWER CONSUMPTION.

Length of serial data	Speed up ratio (n=400)	Speed up ratio (n=625)	Speed up ratio (n=900)
10000	8.94	12.97	17.61
40000	8.60	12.56	17.22
90000	8.33	12.28	16.80
160000	8.07	12.07	16.53
250000	7.95	17.92	16.30
360000	7.79	11.62	16.14
490000	7.64	11.44	16.00
640000	7.04	11.27	15.89

training phase. In fact, an unsupervised process performs the role of front-end data compression. The unsupervised/supervised learning concept relieves the neural network of the burden of trying to associate many dissimilar patterns with the same output. If two groups, which are not alike, happen to end up in the same cluster, the message is that either pattern description is not adequate or vigilance factor is not adequately stringent. Using this concept, the supervised learning process is carried out on the cluster-wise data structure rather than on the entire data set.

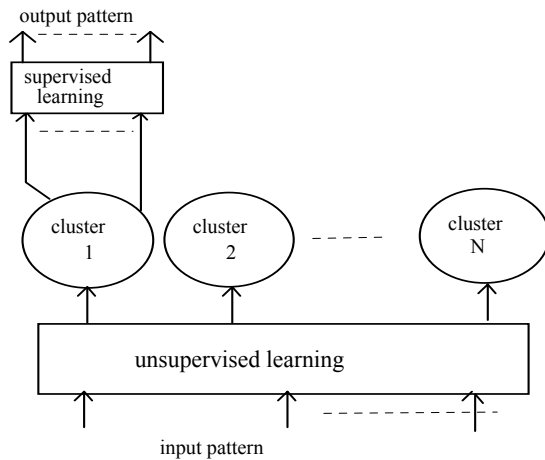


Figure 1. Unsupervised/supervised concept-type1

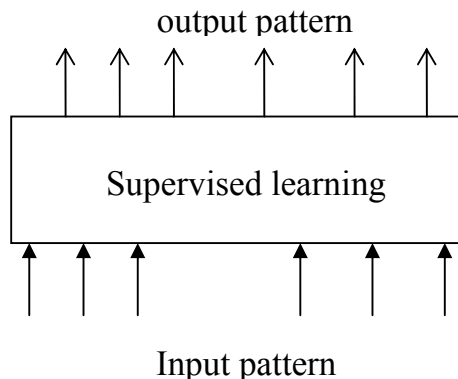


Figure 2. Supervised concept-type

MKNNs consists of N input nodes and M output nodes arranged in a two dimensional grid as shown in Fig.3. The MKNNs has the following purposes [3,5]:

- 1) Clustering of the input space into a finite number of classes represented by the neural networks weight vectors
- 2) Topology preserving mapping of high dimensional input vectors into a lower dimensional surface represented by the location of the neuron on the grid.

MKNNs applies cross correlation in the frequency domain between input data and the weights of neural networks. In MKNNs, unsupervised learning is achieved

in the feature map layer through competition. When an input pattern is presented to the neural network, it computes the activation value for each output node based on present connection weights. The input pattern is said to be mapped to the output node with maximum activation value. After enough input pattern vectors have been presented, input patterns with similar features will be mapped to the same output unit or to output units within a small neighborhood. This unit and its immediate neighbors on the grid are the only units permitted to learn in this pattern presentation by updating their weight vectors. During training, the weight adjustment is proportional to the difference between the input vector and weight vector.

In this paper, MKNNs maps the N inputs x_1, x_2, \dots, x_N ($N=24$) on the M output node y_1, y_2, \dots, y_M . The output node are arranged on a two dimensional grid-like network. Because the clustering of input pattern vectors is self-organized in the learning process, and the ordering of the output node of an input pattern vector is based on that feature, this kind of neural network is called the self-organizing feature map by MKNNs.

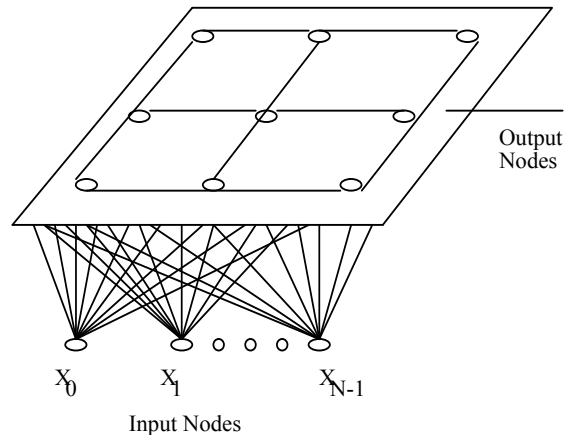


Figure 3. Modified Kohonen's Neural Networks.

Fig. 4 shows the flowchart of the algorithm to produce the self-organizing feature map that can be described as follows:-

1. Read the input patterns $X(1), X(2), \dots, X(P)$.
2. Select the initial values for the connection weights w_{ij} ($i=1, \dots, N, j=1, \dots, M$) and the radius of the neighborhood N_c .
3. The continuous-valued input patterns and the connection weights will give the output node j a continuous activation value which is given by
4. The node j^* with maximum activation value is picked up. The connection weights for that node and all the nodes in the neighborhood defined by N_j are updated using the following equation

$$a_j = \sum_{i=1}^N w_{ij} x_i$$

5. Where, eta is the step size for the updating.
6. Repeat for each input pattern. When all P input patterns have been presented, then one iteration has been completed.

$$w_{i,j \text{ updated}} = w_{ij} + \eta * (x_i - w_{ij})$$

7. Because there are ten iterations sharing one common radius of neighborhood and There are different values of Nc in the training process, then the iteration are repeated until Nc=0.

After training, any input vector stimulates only the neuron whose weight vector is the closest to the input vector in the input space. The weight vectors therefore represent certain averages of disjoint set or classes of input vectors. Further input vectors which are close in the input space stimulate neurons which are close to each other on the grid. Other neurons may not be stimulated by any input vector. Regrouping those neurons may not be stimulated by the same group of the input vectors leads to the concept of neuron clusters represented by output classes. The characteristics of this unsupervised model [3,5] are:

- winner-take-all processing
- unsupervised learning
- lateral connected architecture used for topology preserving clustering and classification

Fig.5 shows the main components of the proposed prediction system.

This paper discussed the data classification using MKNNs for power system.

The elements of the new techniques are:-

- 1 -Historical DB that contains the recording hourly power consumptions, temperatures, wind speed, humidity, atmospheric pressure, from power system, through Data collection module.
- 2- Networks configuration which contains the parameters of MKNNs (number of input nodes, number of output nodes, iterations, number of patterns, and the Files for:- output results, Input Patterns, Input and Output weights, and the learning rate).
- 3- Kohonen's software that implements the self-organization Algorithms.
- 4- Self-organization user Interface As shown in Fig.6.
- 5- Output classes :- represent the day types/classes for the patterns of historical database.
- 6- Supervised software that implement the Linear Vector Quantization Learning Algorithm (LVQLA).
- 7- LVQ user interface that represent a graphic interface for supervised learning that calls the LVQLA.
- 8- FFNN(1) ... FFNN(n):- represent feed-forward neural network for class 1, 2,... N respectively. These

networks are used for prediction process for all day types.

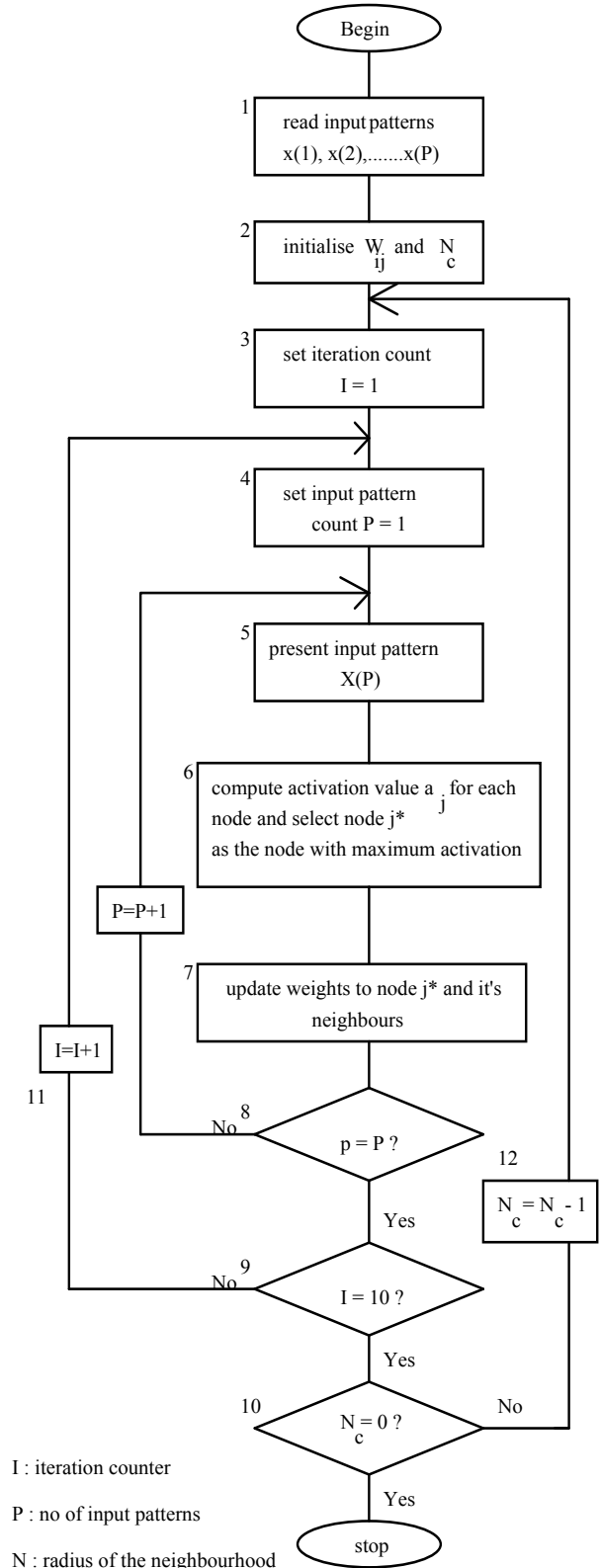


Figure 4. The algorithm of MKNNs.

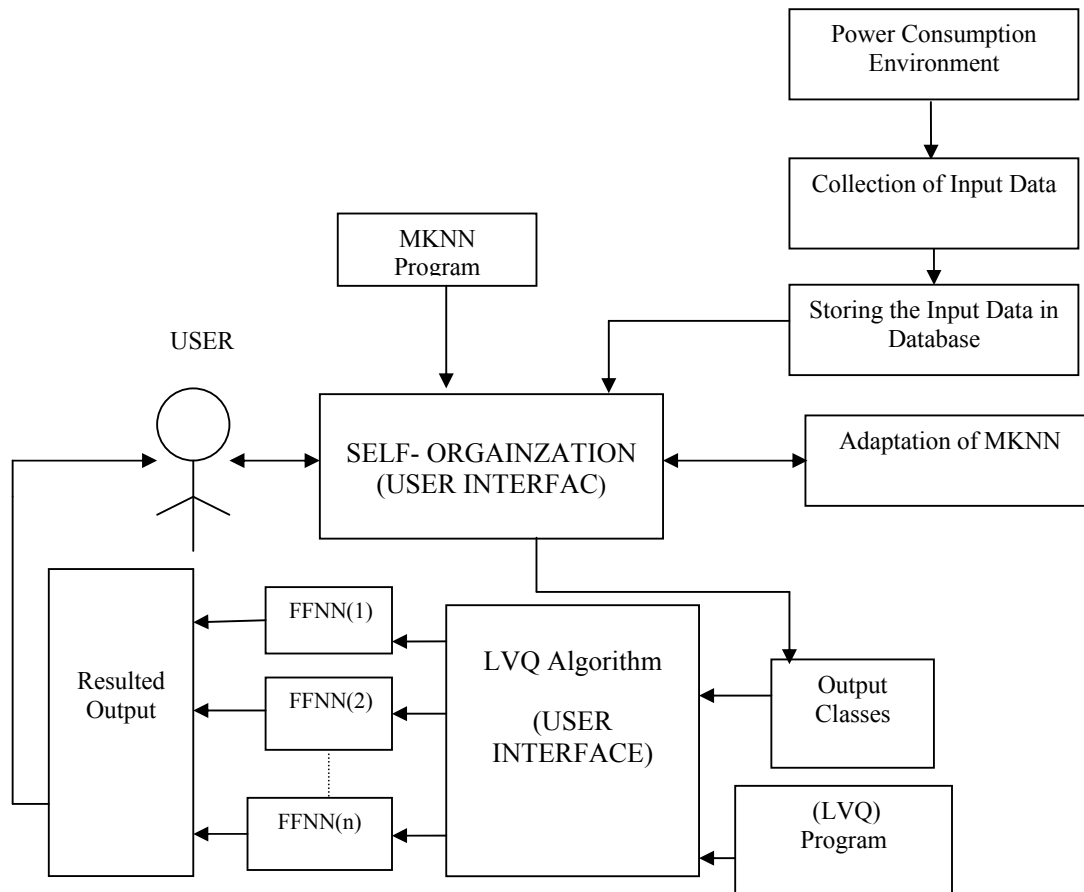


Figure 5. The structure of the prediction system.

V. CONCLUSION

A New hybrid neural model for prediction of power consumption has been presented. The proposed technique integrates the benefits of feedforward and feedback neural networks. It combines the unsupervised and supervised learning concept which classifies input patterns into classes and supervised learning and testing within each formed class. KNNs which uses self-organizing feature map has been modified to identify the day types which are essential to power consumption prediction processes. The inputs patterns of MKNNs are the 24 hourly power consumption patterns. Each input power consumption pattern is mapped to a node on the output plane according to its feature. Similar power consumption patterns are mapped to the same node or neighboring nodes. Examination of the power consumption patterns for each power system, has indicated that four week classes are working days, Thursday, Friday, and Saturday and special events. Each class should be separately used a neural network in the training process. Furthermore, it has been shown that FFNS are very effective in day type classification. The idea of these is to apply cross correlation in the frequency domain between the input data and the weights of neural networks. Moreover, It has been proven that the capability of MKNNs for identifying a new type of power consumption pattern (due to holiday, special event, bad data) before the expert operator is very high. Therefore, the proposed technique can be used as a valuable aid to the system operator in the prediction process.

REFERENCES

- [1] Hazem M. El-Bakry, and Nikos Mastorakis, "A New Fast Forecasting Technique using High Speed Neural Networks," WSEAS Transactions on Signal Processing, Issue 10, vol. 4, October 2008, pp. 573-595.
- [2] Hazem M. El-bakry, and Nikos Mastorakis, "Prediction of Market Price by using Fast Time Delay Neural Networks," Proc. of 10th WSEAS Int. Conf. on Neural Networks (NN'10), Romania, June 13-15, 2010, pp. 230-237.
- [3] Hazem M. El-bakry, and Wael A. Awad, "Fast Forecasting of Stock Market Prices by using New High Speed Time Delay Neural Networks," Waset International Journal of Computer and Information Engineering, vol. 4, no.2., 2010, pp. 138-144.
- [4] Hazem M. El-Bakry, and Nikos Mastorakis, "A New Fast Forecasting Technique using High Speed Neural Networks," 8th WSEAS International Conference on SIGNAL, SPEECH AND IMAGE PROCESSING (SSIP '08), Santander, Cantabria, Spain, September 23-25, 2008, pp. 116-138.
- [5] Hazem M. El-Bakry, "A Novel High Speed Neural Model for Fast Pattern Recognition," Soft Computing Journal, vol. 14, no. 6, 2010, pp. 647-666.
- [6] Hazem M. El-Bakry, "New Faster Normalized Neural Networks for Sub-Matrix Detection using Cross Correlation in the Frequency Domain and Matrix Decomposition," Applied Soft Computing journal, vol. 8, issue 2, March 2008, pp. 1131-1149.
- [7] Hazem M. El-Bakry, "Face detection using fast neural networks and image decomposition," Neurocomputing Journal, vol. 48, 2002, pp. 1039-1046.
- [8] Hazem M. El-Bakry, "Human Iris Detection Using Fast Cooperative Modular Neural Nets and Image Decomposition," Machine Graphics & Vision Journal (MG&V), vol. 11, no. 4, 2002, pp. 498-512.
- [9] Hazem M. El-Bakry, "New Fast Principal Component Analysis For Real-Time Face Detection," Machine Graphics & Vision Journal (MG&V), vol. 18, no.4, 2009, pp. 405-426.
- [10] Hazem M. El-Bakry, "A New Neural Design for Faster Pattern Detection Using Cross Correlation and Matrix Decomposition," Neural World journal, Neural World Journal, 2009, vol. 19, no. 2, pp. 131-164.
- [11] Hazem M. El-Bakry, "An Efficient Algorithm for Pattern Detection using Combined Classifiers and Data Fusion," Information Fusion Journal, vol. 11, 2010, pp. 133-148.
- [12] Hazem M. El-Bakry, "New Fast Time Delay Neural Networks Using Cross Correlation Performed in the Frequency Domain," Neurocomputing Journal, vol. 69, October 2006, pp. 2360-2363.
- [13] Hazem M. El-Bakry, "A New High Speed Neural Model For Character Recognition Using Cross Correlation and Matrix Decomposition," International Journal of Signal Processing, vol.2, no.3, 2005, pp. 183-202.
- [14] Hazem M. El-Bakry, "New High Speed Normalized Neural Networks for Fast Pattern Discovery on Web Pages," International Journal of Computer Science and Network Security, vol.6, No. 2A, February 2006, pp.142-152.
- [15] Hazem M. El-Bakry "Fast Iris Detection for Personal Verification Using Modular Neural Networks," Lecture Notes in Computer Science, Springer, vol. 2206, October 2001, pp. 269-283.
- [16] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Normalized Neural Processors For Pattern Detection Based on Cross Correlation Implemented in the Frequency Domain," Journal of Research and Practice in Information Technology, Vol. 38, No.2, May 2006, pp. 151-170.
- [17] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Time Delay Neural Networks," International Journal of Neural Systems, vol. 15, no.6, December 2005, pp.445-455.
- [18] Hazem M. El-Bakry, and Qiangfu Zhao, "Speeding-up Normalized Neural Networks For Face/Object Detection," Machine Graphics & Vision Journal (MG&V), vol. 14, No.1, 2005, pp. 29-59.
- [19] Hazem M. El-Bakry, and Qiangfu Zhao, "A New Technique for Fast Pattern Recognition Using Normalized Neural Networks," WSEAS Transactions on Information Science and Applications, issue 11, vol. 2, November 2005, pp. 1816-1835.
- [20] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Complex Valued Time Delay Neural Networks," International Journal of Computational Intelligence, vol.2, no.1, pp. 16-26, 2005.
- [21] Hazem M. El-Bakry, and Qiangfu Zhao, "Fast Pattern Detection Using Neural Networks Realized in Frequency Domain," Enformatika Transactions on Engineering, Computing, and Technology, February 25-27, 2005, pp. 89-92.
- [22] Hazem M. El-Bakry, and Qiangfu Zhao, "Face Detection Using Fast Neural Processors and Image Decomposition," International Journal of Computational Intelligence, vol.1, no.4, 2004, pp. 313-316.
- [23] Hazem M. El-bakry, and Nikos Mastorakis, "A New Approach for Prediction by using Integrated Neural Networks," Proc. of Int. Conf., Mexico, Jan. 29-31, 2011, pp. 17-28.
- [24] Hazem M. El-bakry, and Mohamed Hamada "Fast Time Delay Neural Networks for Detecting DNA Coding Regions," Springer, Lecture Notes on Artificial Intelligence (LNAI 5711), 2009, pp. 334-342.
- [25] Hazem M. El-Bakry, and Nikos Mastorakis, "A New Fast Forecasting Technique using High Speed Neural Networks," WSEAS Transactions on Signal Processing, issue 10, vol. 4, October 2008, pp. 573-595.
- [26] Hazem M. El-Bakry, "Automatic Human Face Recognition Using Modular Neural Networks," Machine Graphics & Vision Journal (MG&V), vol. 10, no. 1, 2001, pp. 47-73.
- [27] Hazem M. El-Bakry, and H. Stoyan, "FNNs for Code Detection in Sequential Data Using Neural Networks for Communication Applications," Proc. of the First International Conference on Cybernetics and Information Technologies, Systems and Applications: CITSA 2004, 21-25.
- [28] Hazem M. El-Bakry, and Nikos Mastorakis, "Fast Word Detection in a Speech Using New High Speed Time Delay Neural Networks," WSEAS Transactions on Information Science and Applications, issue 7, vol. 5, July 2009, pp. 261-270.
- [29] Hazem M. El-Bakry, "Fast Virus Detection by using High Speed Time Delay Neural Networks," Journal of Computer Virology, vol.6, no.2, 2010, pp.115-122.

- [30] Hazem M. El-Bakry, and Nikos Mastorakis, "Fast Image Matching on Web Pages," WSEAS Transactions on Signal Processing, issue 4, vol. 5, June 2009, pp. 157-166.
- [31] Hazem M. El-Bakry, and Nikos Mastorakis, "Fast Information Retrieval from Web Pages," WSEAS Transactions on Information Science and Applications, issue 6, vol. 6, June 2009, pp. 1018-1036.
- [32] Hazem M. El-Bakry, and Nikos Mastorakis, "Fast Detection of Specific Information in Voice Signal over Internet Protocol," WSEAS Transactions on Communications, issue 5, vol. 8, May 2009, pp. 483-494.
- [33] Hazem M. El-Bakry, and Ahmed A. Mohammed, "Optimal Document Archiving and Fast Information Retrieval," The International Journal of Computer science, and Engineering, issue 2, vol. 3, July 2009, pp. 108-121.
- [34] Hazem M. El-bakry, and Nikos Mastorakis "Fast Time Delay Neural Networks for Word Detection in a Video Conference," Proc. of European Computing and Computational Intelligence International Conference, Tbilisi, Georgia, June 26-28, 2009, pp. 120-129.
- [35] Hazem M. El-bakry, and Nikos Mastorakis, "Fast Packet Detection by using High Speed Time Delay Neural Networks," Proc. of the 10th WSEAS Int. Conference on Multimedia Systems & Signal Processing, Hangzhou University, China, April 11-13, 2010, pp. 222-227.
- [36] Hazem M. El-Bakry, and Ahmed Atwan, "Improving Quality of Business Networks for Information Systems," The International Journal of Computer, Information, and systems science, and Engineering, issue 3, vol. 3, July 2009, pp. 138-150
- [37] Alaa M. Riad, Hazem M. El-bakry, and Nikos Mastorakis, "Fast Harmonic Current / Voltage Prediction by using High Speed Time Delay Neural Networks, " Proc. of WSEAS International Conference on Communication and Information, Athens, Greece, December 29-31, 2009, pp. 245-272.
- [38] A. M. Riad, and A. K. Mahmoud, "Information System based-on Unsupervised/ Supervised Neural Networks for Daily Electric Load Forecasting process", Egyptian Computer Journal, ISSR, Cairo Univ., 2001.
- [39] A. M. Riad, A. E. M. EL-Misiery, A. A. Soliman and A. K. Mahmoud, "A Hybrid Intelligent System for Forecasting Process", Egyptian Computer Journal, ISSR, Cairo Univ., 2002.
- [40] "Neural Networks -An introduction", http://www.neosciences.com/nn_intro.htm, 2000.
- [41] "Neural Networks –Supervised learning ",http://www.neosciences.com/nn_spvs2.htm, 2000.
- [42] "Neural Networks –Unsupervised learning ",http://www.neosciences.com/nn_unsup.htm, 2000.
- [43] " Artificial Neural Network Models",<http://www-person.usyd.edu.au/~desm/afc-ann.html>, 1999.
- [44] D. W. Patterson, "Artificial neural networks: Theory and applications", Prentice Hall, 1996.
- [45] K. Liu, S. Subbarayan, R. R. Shouults, M. T. Manry. C. Kwan, and F. L. Lewis, "Comparison of very short-term load forecasting techniques", IEEE Trans. PWS, Vol. 11, 1996, pp. 877-882.
- [46] " Neurocomputing for energy forecasting", <http://www.cordis.lu/esprit/scr/results/pages/energy/energy2.htm>, 1999.
- [47] Y. Y.Hsu, and C. C. Yang, " Design of artificial neural networks for short-term load forecasting. Part1: Self-organization feature maps for day type identification", IEE Proc.-C, 1991, pp. 407-413.
- [48] Y. Y. Hsu, and C. C. Yang, " Design of artificial neural networks for short-term load forecasting. Part2 : Multilayer feedforward networks for peak load and valley load forecasting", IEE Proc.-C, 1991, pp. 414-418.
- [49] M. Djukanovic, B. Babic, D. J. Sobajic and Y. H. Pao " Unsupervised/ supervised learning concept for 24-hour load forecasting", IEE Proc.,Vol.140, 1993, pp. 311-318.
- [50] A. G. Tejedor, M. J. Cosculluela, C. Bermejo and R. Montes " A neural system for short term load forecasting based on day-type classification", ISAP-94, France, 1994, Volume 1, pp. 35360.
- [51] R. Satoh, E. Tanaka, and J. Hasegawa " Daily load forecasting using a neural network combined with regression analysis ", ISAP-94, France, 1994, Volume 1, pp. 345- 352.
- [52] "Forecast of Technology",http://www.appliedfutures.com/reports/intelligent_agents.html 2000.
- [53]Stevenson, Howard, ed. *Do lunch or be lunch*. Boston: Harvard Business School Press, 1998
- [54] Flyvbjerg, B., 2008, "Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice." European Planning Studies, vol. 16, no. 1, January, pp. 3-21.

Energy Efficient Cluster Head Election Using Fuzzy Logic In Wireless Sensor Networks

Mostafa Basirnezhad
Department of Computer Engineering
Islamic Azad University of Mashhad
Mashhad, Iran
Basir80@gmail.com

Dr. Masood Niazi Torshiz
Department of Computer Engineering
Islamic Azad University of Mashhad
Mashhad, Iran
Masood.niazi@gmail.com

Abstract— Routing problem is one of important issues in wireless sensor networks and many methods for this purpose is presented. One of these methods is the method of clustering. In this method suitable cluster head can be effective for efficiency and network lifetime. Nodes in the clustering approach can be divided into several groups and each group has a node with the name of cluster head which collect data from the rest nodes of the cluster and leads them to the sink. Although clustering reduces energy consumption, there might be some problem. The main problem is that the most energy consumption is gathered on the cluster head. To overcome this problem in clustering methods, energy consumption should be distributed that could be done by choosing appropriate Cluster head. In this paper a clustering trend based on the clusters reorganization and selecting fuzzy Cluster head is introduced. Fuzzy variables remaining energy of node distance of node to sink and average distance of node to other nodes to select Cluster heads considered. We compare the proposed algorithm terms of energy consumption and network lifetime with LEACH, Gupta and CHEF methods. Simulation results show that the proposed method in comparison with other methods have been improved considerably.

Keywords- increased network lifetime; clustering; fuzzy logic; wireless sensor networks

I. INTRODUCTION

In recent years are technological advances and the telecommunications industry, small electrical and electronic components, leading to construction of small and relatively inexpensive sensors which through a wireless network relate to each other. The networks that wireless sensor networks to be known, to a suitable tool for extracting data from the environment and monitoring environmental events have become and their Application in household, industrial and military day to day is increasing. In wireless sensor network design major issue is limited sensor energy source. On the other hand there because too many sensor in network or lack of access to them, or replace the sensor battery is not practical. Therefore provide ways to optimize energy consumption, which ultimately will increase the network lifetime, is strongly felt [1].

A sensor network consists of many sensor nodes in an environment which is widely distributed and collected data from the environment. Not necessarily located where the sensor nodes are clear. Such properties to the possibility that

we can provide them in dangerous or inaccessible places we leave. Each sensor node contains a processor on your board and instead of sending all the raw data to center or node responsible for information processing and concluded, his first in a series of basic and simple processing on the information gained, performs and will send the data semi-processed [2].

Work on wireless sensor networks in the first began defense and military purposes, but became also many other applications quickly that some of the applications of this technology is in military and security applications, monitor internal and external environments (used in buildings Intelligent, traffic control, detection of natural disasters, agriculture and environmental monitoring), industrial applications and medical applications (health care and surgery) [3].

Other articles seek to be as follows. In second section LEACH algorithm and two fuzzy logic methods for selecting Cluster head use is discussed. The third section main model in wireless sensor networks, we define. In Fourth section, the proposed model is introduced. In fifth Section simulation results and charts shown and in sixth section we express Conclusion.

II. RELATED WORK

One of the introduced and old clustering protocols is LEACH protocol [4] in wireless sensor networks that has been established two Setup phase and steady state phase. Steady state phase in a single data transmission step occurs. Each node in the cluster such as Cluster head is elected. Data collected from member node, before a base station or sink, locally processed data in Cluster head added if any of it is removed and then form a new package will be sent to the base station, the Because energy consumption more than node Cluster head is normal after a while their energy is all. Therefore, the use of clustering in LEACH has been dynamic [5].

This means that after each course and receive operations, in other words after each launch the implementation phase, Cluster head is changed and another node randomly chosen to be as Cluster head. But random selection Cluster head improper distribution may lead them in the network. This means that part of the area covered by two or more are gathered Cluster head while in another area there are no Cluster head.

LEACH clustering algorithm in the model locally and is based on probability models. In this algorithm for each node n Cluster head being a random number between zero and one will choose. If the number is smaller than a threshold $T(n)$, the node for the current round is Cluster head. Threshold is defined by equation (1).

$$T(n) = \begin{cases} \frac{P}{1 - P(r \bmod \frac{1}{P})} & n \in G \\ 0 & n \notin G \end{cases} \quad (1)$$

In equation (1) we have:

P is probability of being Cluster head.

r is current round number.

G is a set of nodes which in $1/p$ before round have not been Cluster head.

When Cluster head were selected, a message to other network nodes as their introduction will send. The nodes, their clusters are selected. This choice is based on the received signal strength. After the time clusters formed, Cluster head a TDMA schedule to create and to each node of a period when the node can send information to offers. This schedule is broadcast in the entire cluster.

During steady state phase, sensors located in the cluster can receive and operate the data to do Cluster head. This information is Cluster head gathering, compression, and then they are submitted to the sink. After a specified time phase network is launching new and Cluster head are elected. In clusters to communicate and to reduce interference nodes belonging to different clusters reject CDMA is used [6].

Although LEACH increases the network lifetime, but in this protocol there are several assumes. LEACH assumes that all nodes have enough energy to have communication with the sink and each node has computational power to support the MAC protocol is different. Therefore applicable to large networks is not. Also assume that nodes always have data to send. Consequently nodes near the data they send a similar. Also unknown is how Cluster head are in the distribution network means may all together in a corner of the network and therefore are there will be sensors that are no Cluster head.

So the idea of dynamic clusters overhead brings a lot. Cluster head such changes and confirmation messages that may be reduced energy. Finally, the protocol assumes that all nodes with a certain amount of energy in each round are elected and assumes that energy consumption Cluster head almost equal with the other nodes [7].

In Gupta method Cluster head at each round by calculating the chance of each node for to be Cluster head described with three fuzzy term node density, energy level and the centrality of each node is selected. In This algorithm central control is in BS that has the network global information and select best Cluster head. BS is many times more powerful than sensor nodes, and has enough memory and power [8].

In Gupta method is consumed energy for all nodes send location information to the BS. Considering wireless sensor networks for development on a geographical area of the main objectives and receive collected information is meaningful. The assumption is that nodes at least are moving, so send location information during setup phase is sufficient. Operation of fuzzy cluster head selected is divided into two stages, each is including a setup phase and steady state phase similar to LEACH algorithm. During the setup phase, Cluster head are determined by fuzzy knowledge process and then clusters are formed. In steady state phase Cluster head given by aggregation are collected and signal processing functions for data compression are performed in a signal. Then the combined signal is sent to the BS. Radio model used in this method is the same model of LEACH [9].

In this method, all nodes based on their chance are compared and node with the highest chances to be selected as Cluster head. Each node in the cluster joins itself to Cluster head and will start sending data. Data transmission phase is similar to steady state phase of LEACH algorithm.

This method increased somewhat network lifetime, but the problem is central selection mechanism, i.e., sink should be all about energy and the distance of the sensor node can collect this information and in accordance with fuzzy logic, to select Cluster head and other problems that is that each round is only one choice Cluster head and are causing many overhead in sink [9].

CHEF method [10] was introduced a mechanism elected Cluster head locally using fuzzy logic. This method on factors that affects the network lifetime defines and according to these factors and IF-Then fuzzy rules, Cluster head is select. In this method, two fuzzy sets of local energy and distance (total distance between node and node which are at r distance.) Is defined and using two sets Cluster head chances for getting a node is calculated. Fig. 1 and equation (2) shows to calculate the local distance.

$$L.D. = d1 + d2 + d3 + d4 + d5 + d6 \quad (2)$$

r mean radius of clusters is desired. Equation (3) shows the calculated r . In this Equation n is number of sensors in wireless sensor network.

$$r = \sqrt{\frac{\text{area}}{\pi \times n \times p}} \quad (3)$$

CHEF method avoids problems of LEACH method and the sensor node that has more energy is being better chance for Cluster head. In addition, fuzzy variable local distance, locally to be selected optimal node that has the most chances as Cluster head because energy consumption in wireless communication depends on distance between two nodes and

node density. According to the local distance and energy, CHEF can choose optimal Cluster head at any time and to increase the network lifetime [11].

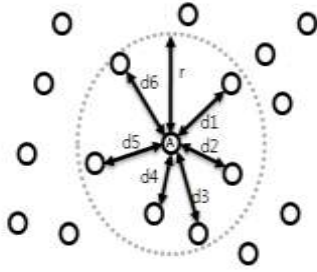


Figure 1. Calculate the local distance

CHEF also overcome the problems of method Gupta and BS will not need to collect fuzzy information of nodes and will cause additional overhead. Furthermore, the method of Gupta choice only one Cluster head at each round.

III. SYSTEM MODEL

In this case the system model is given to each sensor node sends its Cluster head and Cluster head, the collected data aggregation and that they will send to the sink. Some of the assumptions of this method follow [3]:

- Sensor nodes are homogeneous.
- Distance can be by wireless radio signals can be measured.
- Nodes are fixed and no motion.
- All sensor nodes have the same primary energy.

Sink in the center of the sensor network is located.

Equation (4) the amount of energy consumed in a L bit depending on the distance d shows. E_{elec} are amount of energy consumed per bit to run radio transmitter and receiver circuits. D_0 values of the equation (5) are obtained and that the E_{fs} and E_{mp} are amount of energy wasted by the amplifier radio bits [2], [13].

$$E_{tx} = \begin{cases} l * (E_{elec} + E_{fs} * d^2) & d \leq d_0 \\ l * (E_{elec} + E_{mp} * d^4) & d > d_0 \end{cases} \quad (4)$$

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \quad (5)$$

Amount of energy consumed in receiving a L-bit package as equation (6) is calculated.

$$E_{rx} = l * E_{elec} \quad (6)$$

IV. PROPOSED ALGORITHM

In this method sensor nodes randomly in a circular area of radius R are placed. Each node will send position and its remaining energy to the sink located in the center area. Sink based on the percentage of expected Cluster head (p) can share the network a few sectors. The total network is divided into $N * P$ sector that N is number of wireless sensor network node and the angle between the two sectors (θ) are obtained by equation (7).

$$\theta = \frac{2\pi}{N * P} \quad (7)$$

Then each of these sectors according to Fig. 2 lays a cluster with the angle θ and in any cluster selects a Cluster head using fuzzy sets.

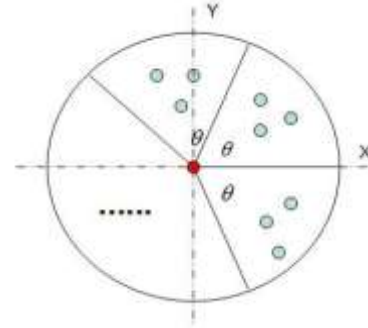


Figure 2. Proposed system model

Select Cluster head in proposed method is doing with defining three fuzzy sets the residual energy of each node, the distance between node to sink and average distance of node to other node. Output fuzzy variable is the chance of node for select Cluster head. In a next stage by adding a fixed amount to θ that we show it with ϕ , the clusters are re-organized and new Cluster head such prior steps are select to the fuzzy methods. With this work nodes that at previous stages have not been selected Cluster head have better chance to be Cluster head. Angle ϕ can be according to the number of sectors such as Fig. 3 was determined. This continues until over first energy sensor node.

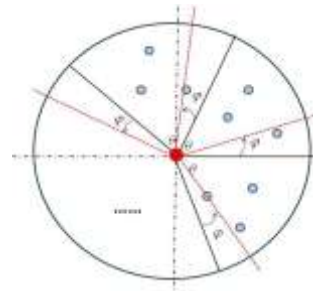


Figure 3. Changing clusters

Input fuzzy variables (residual energy of each node, the distance between node to sink and average distance of node to other node) and output fuzzy variable (the node chance to select Cluster head) are defined as follows:

Fuzzy variable residual energy of each node are divided into three levels Low, Med and High and fuzzy variables distance to sink and average distance node to other node are divided into three levels Near, Med and Far. Output means chance to choose Cluster head are divided into nine levels VeryWeak, Weak, LittleWeak, LowMed, Med, HighMed, LittleStrong, Strong, and VeryStrong. Membership functions of residual energy, distance to sink, the average distance to other nodes and the chances can see in Fig. 4, 5, 6 and 7 respectively.

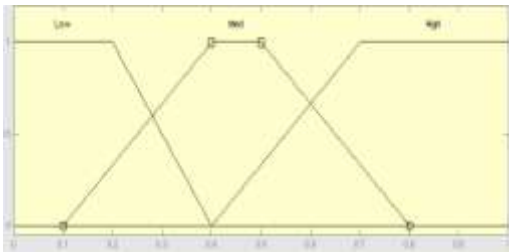


Figure 4. Membership functions of residual energy of node

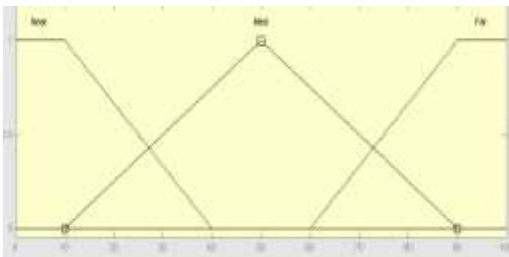


Figure 5. Membership functions of distance node to sink



Figure 6. Membership functions of average distance of node to other node

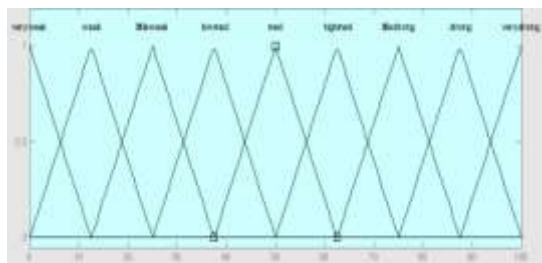


Figure 7. Membership functions of chance of node

TABLE I. FUZZY RULES OF PROPOSED ALGORITHM

ID	Energy	Distance	AverageDistance	Chance
1	Low	Far	Far	VeryWeak
2	Low	Far	Med	VeryWeak
3	Low	Far	Near	Weak
4	Low	Med	Far	VeryWeak
5	Low	Med	Med	Weak
6	Low	Med	Near	LittleWeak
7	Low	Near	Far	Weak
8	Low	Near	Med	LittleWeak
9	Low	Near	Near	LowMed
10	Med	Far	Far	LowMed
11	Med	Far	Med	LowMed
12	Med	Far	Near	Med
13	Med	Med	Far	LowMed
14	Med	Med	Med	Med
15	Med	Med	Near	Med
16	Med	Near	Far	Med
17	Med	Near	Med	HighMed
18	Med	Near	Near	HighMed
19	High	Far	Far	LittleStrong
20	High	Far	Med	LittleStrong
21	High	Far	Near	Strong
22	High	Med	Far	LittleStrong
23	High	Med	Med	Strong
24	High	Med	Near	Strong
25	High	Near	Far	Strong
26	High	Near	Med	VeryStrong
27	High	Near	Near	VeryStrong

According to the fuzzy input and output variables, fuzzy rules base is including 27 law that is shown in Table (1). Our fuzzy logic control model includes a fuzzifier, fuzzy rules, fuzzy inference engine and is a de-fuzzifier. In this paper, the most famous model named fuzzy inference technique is used Mamdani. For the output membership function is expressed center of gravity method is used as equation (8).

$$COG = \frac{\sum \mu_A(x) * x}{\sum \mu_A(x)} \quad (8)$$

In equation (8) $\mu_A(x)$ is Membership function of fuzzy set A.

V. EVALUATION AND SIMULATION

The proposed algorithm was assessed in MATLAB. For simulation a wireless sensor network with 100 nodes in a circular environment with radius 100m is considered. Nodes distribute randomly in the environment with same energy of one joule and the sink were located in the center region is assumed that no sink in terms of energy restriction and all processing is done in the sink. Position of each node to determine the distance to the sink and angular node that has the X axis in polar coordinates is considered. We compare the proposed algorithm with the LEACH, Gupta and CHEF algorithm of network lifetime. Here the network lifetime when the first network node is lost we have considered. The Fig. 8 shows simulation result after 200 stages. As in Fig. 8 can be seen, the network lifetime compared to other methods has increased. This increase is due to in optimal selection cluster head in each cluster and the appropriate distribution of clusters over the simulation environment.

Network residual energy in each round is good scale to measure performance of algorithm. Lower slope in the diagram of energy balance and better distribution of appropriate energy nodes will specify. The Fig. 9 shows compared the rate of energy consumption in the four methods mentioned above. The Fig. 9 is determined by the network residual energy in the proposed method more than the three other methods.

Distance between Cluster head is one of the important parameters in clustering. In this algorithm order to the region is equal and in each region one Cluster head are selected from rally Cluster head in the near and from distance falling between cluster heads prevents and increase network lifetime and energy consumption will be better.

Fig. 10 shows energy consumption in the proposed algorithm compared with the three other methods mentioned. As received from the chart increase rate of energy consumption in the proposed method than the other three methods is reduced. This reduction of energy consumption due to right distribution of cluster Heads in the environment and the use of fuzzy logic is to select Cluster head.

VI. CONCLUSION

Optimize energy consumption in wireless sensor networks is very important, so that the optimal energy consumption lead to increase network lifetime. In this paper, using the fuzzy as sink run in Central, we have determined place of Cluster heads and we have the minimum energy consumption in the network. Also, with reorganization of clusters we could increase wireless sensor network lifetime.

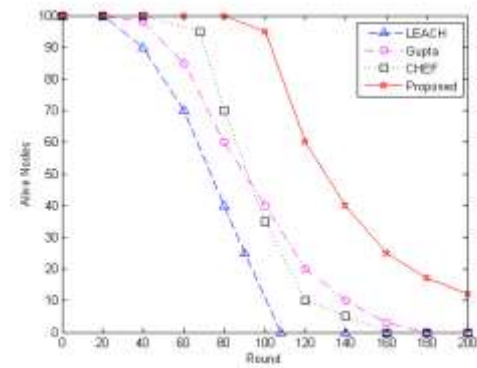


Figure 8. Compare network lifetime for 100 nodes

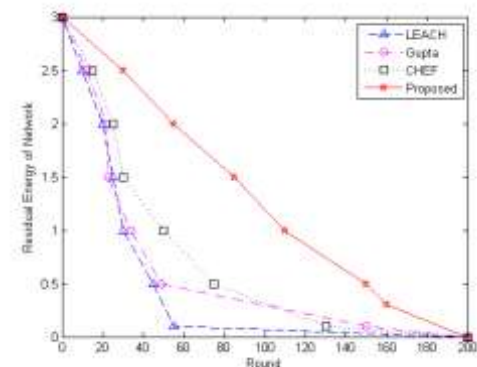


Figure 9. Network residual energy

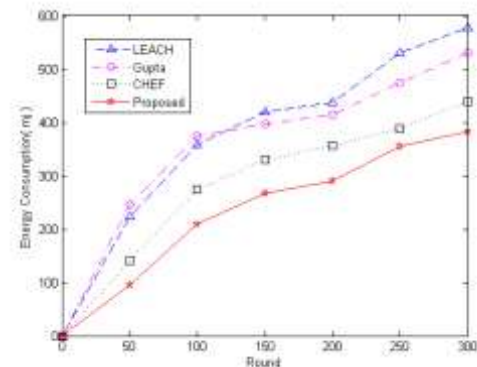


Figure 10. Energy consumption rate

References

- [1] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660-670, October 2002.
- [2] S.H. Park, J.S. Cho, S.M. Jung, Y.J. Han, and T.M. Chung, "Architecture of Context-Aware Integrated Security Management Systems for Smart Home Environment", *APNOMS2007, LNCS 4773*, pp. 543-546, October 2007.
- [3] K. Akkaya, M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks," in *the Elsevier Ad Hoc Network Journal*, Vol.3/3, pp. 325-349, 2005.
- [4] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocols for wireless microsensor

- networks” Proceedings of the Hawaii International Conference on Systems Sciences, Jan. 2000.
- [5] A.Khadem Zadeh, A.Hosseinalipour, S. Babaie "New Clustering protocol for increasing Wireless Sensor Networks Lifetime" International Journal of Computer and Network Security, Vol. 2, No. 1, January 2010.
- [6] Y.J. Han, S.H. Park, J.H. Eom, and T.M. Chung, "Energy-Efficient Distance Based Clustering Routing Scheme for Wireless Sensor Networks", ICCSA2007, LNCS 4706, Part II, pp. 195-206, 2007.
- [7] Namhoon Kim, Jongman Heo, Hyung Seok Kim, Wook Hyun Kwon "Reconfiguration of cluster heads for load balancing in wireless sensor networks" Elsevier, Computer Communications 31 (2008) 153–159.
- [8] I.Gupta, D. Riordan and S. Sampalli, "Cluster-head Election using Fuzzy Logic for Wireless Sensor Networks," In Proceedings of the 3rd Annual Communication Networks and Services Research Conference, pp.255-260, May 2005.
- [9] M.R. Minhas, S. Gopalakrishnan and V. Leung, "Fuzzy Algorithms for Maximum Lifetime Routing in Wireless Sensor Networks", Global Telecommunications Conference, 2008, pp. 1-6.
- [10] J. Myoung Kim, S. Park, Y. Han and T. Chung, "CHEF: Cluster Head Election mechanism using Fuzzy logic in Wireless Sensor Networks," 10th International Conference on Advanced Communication Technology (ICACT), pp.654-659, February 2008.
- [11] Kumar O, Aseri TC and Patel RB, "EEHC : energy efficient heterogeneous clustered scheme for wireless sensor networks," Computer Communication, 32(4), 2009, p p. 662-667.
- [12] Israr N and A wan I, "Coverage based inter cluster communication for load balancing in heterogeneous wireless sensor networks," Telecommunication. Sits, 38(3-4), 2008, pp. 12 1-132.
- [13] Haosong G., Gang L.,Younghwan Y.," A Partition-Based Centralized LEACH Algorithm for Wireless Sensor Networks using Solar Energy". in Proceedings of ICHIT2009.
- [14] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, Andrea Passarella, Energy Conservation in Wireless Sensor Networks: a Survey, Ad Hoc Networks, 2009.

Effective Formal Procedure of Alternate Routings in MANET Improving Quality of Service

Shakeel Ahmed and A. K. Ramani

School of Computer Science and Information Technology
Devi Ahilya University
Indore, INDIA
Emails: {shakeel_rahi, ramaniak}@yahoo.com

Nazir Ahmad Zafar

Department of Computer Science
King Faisal University
Hofuf, SAUDI ARABIA
Email: nazafar@kfu.edu.sa

Abstract—Various protocols are proposed in Mobile Ad Hoc Networks (MANET) for finding the best route from source to destination however, there doesn't exist any real protocol considering all the parameters needed to attain acceptable level of Quality of Service (QoS) for real-time interactive applications. Consequently this is an open research problem and has become a challenging issue for today's network services. Due to mobility of the nodes in MANET the topology changes frequently, the nodes are self organized, lack centralized control and move freely. Many of the routing protocols proposed maintain only one route to the specified node and do not consider important QoS parameters like bandwidth, packet type and the route is rediscovered in these protocols by the source node when the earlier route fails. In this paper, we have proposed a new Formal approach of multipath node-disjoint routings based on AODV protocol which uses two node-disjoint routes between source and destination to improve QoS. In the proposed approach mobility of nodes, bandwidth requirement, and battery life are considered in route discovery process that can significantly reduce end-to-end delay and increase packet delivery ratio. Dynamic graphs are used to model the network topology since the nodes of MANET are not fixed. Investigation and analysis of the proposed routing protocol is carried out by employing formal techniques. Z notation is used to describe the formal model because of its abstract and rigorous power of specifying the properties. The specification is analyzed and validated using Z Eves tool.

Keywords—Ad Hoc networks; Alternate route; QoS; Formal methods; Z notation; Validation

I. INTRODUCTION

Mobile Ad Hoc networks are characterized by dynamic topology due to node mobility, limited bandwidth and power of battery. In this situation, it is essential to find the route with better Quality of Service (QoS) considering low overheads, dynamic conditions, routing with maximal throughput and minimal control overhead. Several on-demand routing protocols are proposed for Ad Hoc networks in which nodes build and maintain routes as needed, and route discovery process is initiated whenever a node needs a route to a particular destination. Examples of such protocols include the Dynamic Source Routing (DSR) [1], Ad Hoc On-Demand Distance Vector (AODV) protocol [2], Temporally Ordered Routing Algorithm (TORA) [3].

Several performance studies [4, 5, 6] of Ad Hoc networks have shown that on-demand protocols incur lower routing overheads compare to the proactive counterparts. It is a well

known fact that the routing consumes both channel bandwidth as well as the battery power of nodes for communication or processing which leads to poor QoS. In single-path On-demand protocols routing of data takes place over a single-path which is discovered during the route discovery process of the routing algorithm. In these routing protocols, route discovery may be performed after route fails. Therefore, data transmission will be stopped while the new route is established causing data transmission delay. On the other hand, multipath routing establishes multiple routes between source and destination nodes. For fault tolerance, even if one route fails, source node can maintain connections by using other routes. Hence multiple routing protocols can reduce data transmission failures and delay times that are caused by route disconnection. To perform the route maintenance process before all routes fail, the network must always maintain multiple routes and the selection criterion of the specific path depends on the routing algorithm employed. A major restrain factor is the control traffic overhead that is incurred as a result of the discovery and subsequent maintenance of multiple routes. However, multipath extensions to on-demand routing protocols exist that keep the low overhead of control.

Having multiple paths in Ad Hoc networks is beneficial because wireless networks are prone to route breaks resulting node mobility, fading environment, signal interference, high error rate, and packet collisions. It is also important to generate alternate routes for the primary route without propagating more control messages than building only single route. Minimization of the number of packet transmissions is critical in Ad Hoc networks with limited bandwidth and battery life.

In this paper, Alternate Route Quality of Service (ARQoS) [7] model is analyzed, verified and validated using formal method to improve QoS in on-demand routing protocols by providing alternate paths for the single-path. Selection of alternate route will be carried out by the protocol itself when the primary route fails.

Most of the proposed protocols are focused on simulation and few implementations are proposed in which environments had no more than a dozen of nodes. It is also observed that for very simple protocols different simulators can produce immensely different results. By using model checking one can present performance measures that are difficult if not impossible to obtain by simulation, and the results obtained from simulations are often not realistic as they have not been validated against empirical data. Simulations do not really

improve the development process. [8]. Graph theory has much of its applications in the area of parallel and distributed algorithms and is an effective tool for modeling and visualizing the communication networks. Graph theory does not have much computer tool support for verifying and validating the systems. Formal techniques are best approaches for specification and proving the computerized models. In this research, formal methods in terms of Z notation [9] are used by linking with graph theory for describing multiple routes by considering QoS parameters and updating the routing table. Z notation is used because of abstraction and encapsulation of objects for further enhancement of the description of the system. Rest of the paper is organized as follows.

In section 2, it is provided an outline of the related work. Section 3 presents an introduction to formal methods. In section 4, an introduction to AODV protocol is given. In section 5, formal specification of ARQoS is described. Finally, concluding remarks are given in section 6.

II. RELATED WORK

There are three main categories of Ad Hoc routing protocols: Proactive (table-driven), Reactive (On-demand) and Hybrid [10]. Proactive protocols build their routing tables continuously by broadcasting periodic routing updates through the network; reactive protocols build their routing table's on-demand and have no prior knowledge of the routes they will take to get to a particular destination. Hybrid protocols create reactive routing zones interconnected by proactive routing links and usually adapt their routing strategy to the amount of mobility in the network [11].

A number of solutions for multipath routing in Ad Hoc networks are based on AODV (Ad Hoc On-demand Distance Vector) [12]. Some of the examples are: AOMDV (Ad Hoc On-demand Multiple Distance Vector) [13], AODVM (Ad Hoc On-demand Distance Vector Multipath) [14], AODV-BR (Ad Hoc On-demand Distance Vector Backup Route) [15] and MP-AODV (Multipath- Ad Hoc On-demand Distance Vector) [16], AODV-ABR (Ad Hoc On-demand Multiple Distance Vector- Adaptive backup routing) and AODV-ABL (Ad Hoc On-demand Multiple Distance Vector - Adaptive backup with local repair routing) [17].

The AOMDV protocol establishes loop-free link-disjoint paths in the network. When intermediate nodes receive the RREQ packet from the source node, AOMDV stores all RREQ packets, unlike conventional AODV, which discards duplicates. In this way, each node maintains a first hop-list where information from additional field called first hop in RREQ packet indicates the neighbor node of the source nodes. If first hop of received RREQ packet is duplicated from its own first hop-list, the RREQ packet is discarded. On the other hand, the RREQ packet is not duplicated from previous RREQ packets. Then the node updates the next hop, hop count and advertised-hop count in routing table. At the destination, RREP packets are sent from each received RREQ packet. The multiple routes are made by RREP packets that follow the reverse routes which have already been setup in the intermediate nodes as discussed in [13].

For the AODVM protocol, intermediate nodes are not allowed to send a RREP packet directly to the source node. Further, intermediate nodes do not discard the duplicate RREQ packets. But the intermediate nodes record all received RREQ packets in routing table. The destination node sends an RREP for all the received RREQ packets. An intermediate node forwards a received RREP packet to the neighbor in the routing table. Whenever a node overhears one of its neighbors broadcasting RREP packet, it removes that neighbor from its routing table, because nodes cannot participate in more than one route [14].

In AODV-BR the authors propose a scheme to calculate alternate paths such that when a link failure occurs, the intermediate node searches for an alternate route to circumvent the broken link. The basic assumption made in this protocol is that all the nodes are in promiscuous mode and that they can overhear every transmission within their range [15]. The above mentioned protocols have a number of limitations. First, none of the protocols have considered the QoS parameters including constant mobility of the nodes and an introduction of new node coming within the transmission range. The newly added may have an optimal path to destination is not considered. The main drawback of AODV-BR protocol is that the alternate routes that are computed during route discovery are not maintained during the course of data transfer and in this way QoS is not met. Thus the routes could become stale and outdated by the time they are actually utilized. Moreover, the utilization of promiscuous mode greatly increases the power consumption of each node.

III. FORMAL METHODS

Formal methods are extensively used in a variety of areas including software engineering, modeling and simulation, verifying network protocols, designing and development of parallel and distributed systems, model checking, theorem proving and for checking hardware systems. Hence Formal methods are best choice for modeling of mobile Ad Hoc networks due to its distributive nature and because of much component of software as compared to its counterpart hardware. An important aspect of a wireless networks is that nodes use multi-hop communication on an unreliable medium, further the network is subject to dynamic changes and environmental interferences therefore algorithms and protocols should be used for analysis, writing formal specification and producing refinements [18].

A formal specification is a description that is abstract, precise and in a sense is complete. The abstraction allows a human reader to understand the big picture; the precision forces ambiguities to be questioned and removed; and the completeness means all aspects of behavior are described [18]. Secondly, the formality of the description allows us to carry out rigorous analysis. By looking at a single description one can determine useful properties such as consistency or deadlock-freedom [19]. By writing different descriptions from different viewpoints one can determine important properties such as satisfaction of high level requirements or correctness of a proposed design

Z notation [20, 21] is a model-based approach which is a strongly typed, mathematical specification language, not an executable notation and it cannot be interpreted or compiled into a running program. There are few tools for checking Z texts for syntax and type errors in much the same way that a compiler checks code in an executable programming language. In Z notation, schemas are used which are small pieces for decomposing a specification into manageable components. The schema is the feature that distinguishes Z from other formal notations. In Z schemas are used to describe both static and dynamic aspects of a system [22]. Z specification enables to produce a model that is unambiguous, verifiable and traceable. Z is more mature and has an ISO standard [23].

In Ad Hoc networks nodes are free to move causing changes in the network topology and highly dynamic. This dynamic nature increases the complexity of the algorithms designed for Ad Hoc networks and the verification of these algorithms is a difficult error-prone task that requires much effort. Formal methods have a lot to offer where mobility of the nodes can be modeled from a complex system to mathematical entities resulting in a rigorous model. By using these techniques it is possible to model and verify the mobility of nodes in a more thorough and detailed fashion than the empirical testing and simulation techniques.

IV. PROTOCOL CONSTRUCTION

In this section, we present the procedure details of the ARQoS protocol. Since the purpose of our study is to improve the performance of existing on-demand protocols specifically of AODV hence suggested modification for improving the QoS is also introduced.

A. Route Construction and Route Reply

In ARQoS the original AODV is extended. When a source needs to initiate a data session to a destination but does not have any route information, it searches a route by flooding a Route Request (RREQ) packet. During the searching process the node considers QoS parameters: mobility, bandwidth, and battery life of the nodes. The steps in Route Discovery process are listed below:

- The source node first calculates the bandwidth needed and examine the links between itself and neighbor nodes. If there is enough available bandwidth, the source node generates a RREQ packet, and sets up a routing table for this data packet and broadcast the RREQ packet.
- Each RREQ packet has a unique identifier so that nodes can detect and drop duplicate packets. An intermediate node, upon receiving a non-duplicate RREQ, records the previous hop and the source node information in its route table, process is called backward learning.
- An intermediate node receiving a RREQ examines the links of the neighbor nodes. If the required bandwidth is available the RREQ packet is rebroadcast and a reverse route to the source node is sets up. If there is

enough available bandwidth till the destination receives the RREQ packet.

- When there are more than one nodes meeting the need of bandwidth, the source node will choose the best route based on the delay as the primary route and next route meet the QoS requirements as alternate route.
- During the route request process the primary and the alternate routes are established considering only the nodes which have met the QoS requirement for the data to be transmitted on the route established.
- Considering mobility of nodes if the primary route fails, alternate route will be considered as primary route. And a secondary alternate route will be found by the source node to backup for the alternate route which may cause a failure of communication.

B. An Example

In Figure 1, primary route and an alternate backup route are represented based on AODV routing protocol. When a link failure occurs, the source node will use the alternate route to transmit the data. The protocol assumes that both the source node and the destination for primary and alternate routes are same but intermediate nodes for both the routes are disjoint. When the alternate route fails, ARQoS will have another route from source to destination which is a substitute for the alternate route.

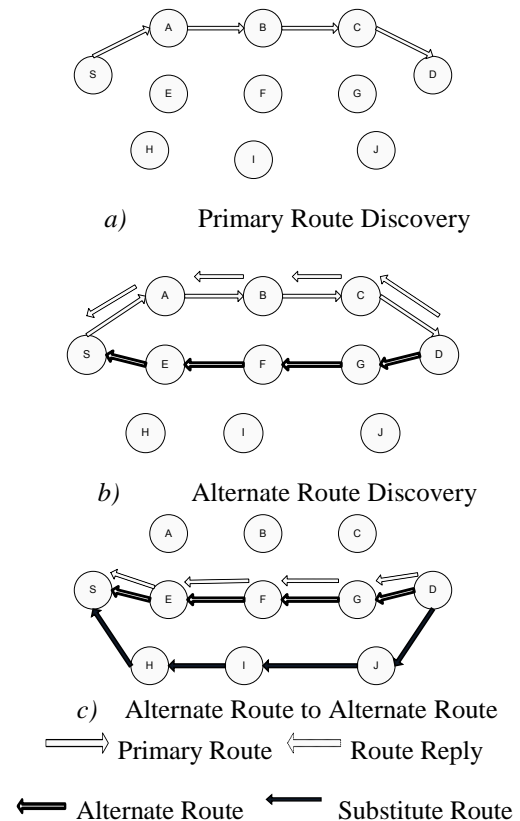


Figure 1. Multiple Route Construction

In AODV, a route is timed out when it is not used and updated for certain duration of time. We use the same technique for timing out alternate routes. Nodes that provide alternate routes overhear data packets and if the packet was transmitted by the next hop to the destination as indicated in their alternate route table, they update the route. If an alternate route is not updated during the timeout interval, the node is removed from the route stored in the routing table.

V. FORMAL ANALYSIS

In this section, formal analysis of the routing table for both primary and alternate routes is presented. The network management for the routing table for ARQoS on-demand distance vector routing protocol is also presented. Initially, formal definitions of basic data types are described and then moving objects and network needed for Ad Hoc network are defined. Finally, single path routing table, alternate path routing table, disjoint routings, and routing table management are described.

A. Formal Model of ARQoS Network

An interconnected collection of objects that help and allow users to share information and resources is termed as communication network. A mobile Ad hoc network is a collection of self-configuring objects inter-connected by wireless links and devices which are free to move in any direction in the domain. This network might be a part of another larger network of communicating objects. In this paper, the communication network is defined by a graph relation where the moving objects are considered as nodes and communication links are assumed as edges of the graph. The graph relation is not static or fixed on the other hand it is dynamic, i.e., its any two nodes may be connected at one time while might be disconnected at another time. In formal definition, the identifier of a moving object is denoted by *Node* as given below. Four types of nodes, i.e., source, destination, internal and nil are assumed which will be needed to analyze and search the route for data transmission. The power of battery is considered and is assumed as dead, low or normal denoted by *Dead*, *Low* and *Normal* respectively.

[Node];

Power ::= Dead | Low | Normal

Type ::= Source | Destination | Internal | Nil

The above types are assumed as sets types in which we do not impose any restriction upon number of elements in a set and as a consequent a high order of abstraction is supposed. Moreover, we do not insist upon any effectual procedure to decide about an arbitrary element if it is a member of the given set in Z notation. Consequently, the *Node* is a set of nodes over which we cannot define an operation of cardinality to know the number of elements. Similarly, the complement and subset operations are not well-defined over sets in Z notation. The sets *Power* and *Type* are assumed as free types in which at one time only one value is assumed.

The moving object of the network is defined as a schema given below which is denoted by *Object* and has four components namely, identification (*id*), type (*type*), battery (*battery*) and set of the neighbors (*neighbors*). The set of neighbors is assumed as a finite power set of *Node* which is defined as an abstract data type. The type of object is considered because it might be a source, destination or an internal node. Further, the node is given a *Nil* value if it is not part of any route stored in the routing table.

Object

id: Node
type: Type
battery: Power
neighbours: \mathbb{F} Node

The possibility of communication between two objects of the graph is defined as an edge which is described by the schema *Connectivity* given below. It consists of three components: *connection*, *trans* and *weight*. The first one is used to define a link between two nodes of the graph. The second component is used to represent the transmission between both the nodes defining edge. The transmission is only possible if the nodes are in range, if nodes are out of range transmission is not possible. To represent the transmission two states are assumed which are denoted by *InRange* and *OutOfRange*. And the last one component *weight* is used to represent the time needed a node to communicate with the other, and is an alternative way of representing bandwidth. Because an object cannot communicate to itself therefore both elements defining connectivity cannot be the same elements.

Transmission ::= InRange | OutRange

Connectivity

connection: Object \times Object
trans: Transmission
weight: \mathbb{N}

connection . 1 . *id* \neq *connection* . 2 . *id*

Object level communication is extended to define the entire network. We have supposed that if there is an edge between two objects then communication is possible. Since any two objects in the network can communicate and hence status of the communication is defined above representing in range or out of range in the definition of connectivity. The formal specification of the network is described by the schema *Network* given below consisting of two components which are *objects* and *connections*. The variable, *objects*, is a collection of nodes of the graph defined as a finite power set of *object*. And the second variable, *connections*, is a finite power set of *Connectivity* used to represent the set of all the edges. The invariants are defined in the predicate part of the schema

Network

objects: F *Object*
connections: F *Connectivity*

$\forall con: Connectivity \mid con \in connections$
• $\exists o1, o2: Object \mid o1 \in objects \wedge o2 \in objects$
• $con.connection = (o1, o2)$
 $\forall o1, o2: Object \mid o1 \in objects \wedge o2 \in objects$
• $\exists con: Connectivity \mid con \in connections$
• $con.connection = (o1, o2) \Rightarrow o1.id \neq o2.id$
 $\forall o1, o2: Object \mid o1 \in objects \wedge o2 \in objects$
• $\exists con1: Connectivity \mid con1 \in connections$
• $con1.connection = (o1, o2) \Rightarrow (\exists con2: Connectivity \mid con2 \in connections \cdot con2.connection = (o2, o1))$

Invariants: (i) For any communication link, there must be two objects which can communication to each other. (ii) For any two objects in the graph relation making an edge, the objects identifiers are different. (iii) It is supposed that if an object can communicate to another then vice versa is also possible that is the graph is a symmetric relation.

In mobile Ad hoc networks, if a node is connected with another node at one time it might be disconnected at the other time. It means the communication is possible only if the nodes are connected. In our model, we have supposed that communication is made if the link between the nodes is active that is nodes are in range to each other. The formal specification of mobile Ad hoc network is described below based on the definition of network.

AdhocNetwork

Network

$\forall o1, o2: Object \mid o1 \in objects \wedge o2 \in objects$
• $\exists con: Connectivity \mid con \in connections$
• $con.connection = (o1, o2)$
 $\Rightarrow con.trans = InRange$
 $\forall o1, o2: Object \mid o1 \in objects \wedge o2 \in objects$
• $\forall con: Connectivity \mid con \in connections$
• $con.connection \neq (o1, o2)$
 $\Rightarrow con.trans = OutRange$

Invariants: (i) For any two objects in the graph relation there must be an edge connecting it which is active or passive. (ii) If objects are out of range to each other the communication is not possible.

The history of routes is stored in the routing table which is defined by the schema *SPRTTable* consisting of two variables. The first one component *AdhocNetwork* is defined above as ad hoc network and the second one is set of primary routes stored in the routing table. The variable *routes* is a collection of paths whereas each path is a well-defined sequence of nodes in the network.

SPRTTable

AdhocNetwork
routes: F (seq *Node*)

$\forall route: seq Node \mid route \in routes$
• $ran route \subseteq \{ o: Object \mid o \in objects \cdot o.id \}$
 $\forall route: seq Node \mid route \in routes \wedge \# route > 1$
• $\forall i: \mathbb{N} \mid i \in 1.. \# route - 1$
• $\exists con: Connectivity \mid con \in connections$
• $(route.i, route.(i+1))$
 $= (con.connection.1.id, con.connection.2.id)$
 $\wedge con.trans = InRange$
 $\forall route: seq Node \mid route \in routes \wedge \# route > 1$
• $\forall i: \mathbb{N} \mid i \in 1.. \# route$
• $\exists object: Object \mid object \in objects$
• $object.id = route.i \wedge object.battery \neq Dead$
 $\forall route: seq Node \mid route \in routes$
• $\# route \geq 1 \wedge (\exists o: Object \mid o \in objects \cdot (o.id = route.1 \wedge o.type = Source)) \wedge (\exists o: Object \mid o \in objects$
• $(o.id = route(\# route) \wedge o.type = Destination))$
 $\wedge (\forall i: \mathbb{N} \mid i \in 2.. \# route - 1$
• $(\exists o: Object \mid o \in objects$
• $(o.id = route(i) \Rightarrow o.type = Internal)))$

Invariants: (i) In this property, it is stated that a route must be a path whose all nodes are objects of the network. (ii) The connectivity of nodes in the route is checked by relating it with edges of graph relation of the ad hoc network. (iii) A node can not be part of the route if its battery is dead. (iv) It is verified that for every route the first node is a source and the last one is a destination. All others are the internal nodes.

Two routes are maintained, at a time, namely primary and alternate if exist. When primary link fails the source node will use the alternate route to transmit the data. If the alternate route fails it will search another route from source to destination, it may be the same as the primary route which was broken. The set of alternate routes is defined in the schema *APRTTable* given below. The same type of components and properties are defined in the alternate routing table as were defined in the primary routing table for well defining the routes. Further, a route is timed out when it is not used and updated for a particular duration of time. If an alternate route is not updated during the timeout interval, the node removes the route from the routing table.

In below, both the primary and alternate routes are defined in separate routing tables. Now we can define a complete routing table based on the primary and alternate routing tables defined above by using the schema *DisjointRoutings*. The relationship between both the routing tables is established. For example, for every primary route there must be an alternate route if exists. The both routes are disjoint. The alternate route is the most appropriate route except the primary route in the ad hoc network. All such properties are defined in the predicate part of the schema.

APRTable

AdhocNetwork

alternatives: F (seq Node)

```
∀route: seq Node | route ∈ alternatives
  • ran route ⊆ { o: Object | o ∈ objects • o . id }
∀route: seq Node | route ∈ alternatives ∧ # route > 1
  • ∀i: ℕ | i ∈ 1 .. # route - 1 • ∃con: Connectivity | con ∈
    connections • (route i, route (i + 1))
    = (con . connection . 1 . id, con . connection . 2 . id)
  ∧ con . trans = InRange
∀route: seq Node | route ∈ alternatives ∧
  # route > 1 • ∀i: ℕ | i ∈ 1 .. # route • ∃object: Object | object ∈
  objects • object . id = route i ∧ object . battery ≠ Dead
∀route: seq Node | route ∈ alternatives | # route ≥ 1
  ∧ (∃o: Object | o ∈ objects • (o . id = route 1 ∧ o . type =
  Source)) ∧ (∃o: Object | o ∈ objects
  • (o . id = route (# route) ∧ o . type = Destination))
  ∧ (∀i: ℕ | i ∈ 2 .. # route - 1 • (∃o: Object | o ∈ objects
  • (o . id = route (# route) ⇒ o . type = Internal)))
```

Invariants: (i) An alternate route must be a well defined path whose all nodes are objects of the network. (ii) Any two consecutives nodes in the alternate route must be an edge of the ad hoc network. (iii) A node can not be included in an alternate route if its battery is dead. (iv) For every alternate route the first node is source and the last one is a destination, all others are internal nodes.

DisjointRoutings

SPRTable

APRTable

```
# routes = # alternatives
∀ route, alter: seq Node
  | route ∈ routes ∧ # route > 1 ∧ alter ∈ alternatives ∧ #
  alter > 1 • route 1 = alter 1 ∧ route (# route) = alter (# alter)
∀ route, alter: seq Node
  | route ∈ routes ∧ # route > 1 ∧ alter ∈ alternatives ∧ #
  alter > 1 • (∀i: ℕ | i ∈ 2 .. # route - 1
  • (∀j: ℕ | j ∈ 2 .. # alter - 1 • route i ≠ alter j))
  ∧ (∀i: ℕ | i ∈ 2 .. # alter - 1
  • (∀j: ℕ | j ∈ 2 .. # route - 1 • alter i ≠ route j))
∀ route: seq Node | route ∈ routes ∧ # route > 1
  • ∀i: ℕ | i ∈ 1 .. # route - 1
  • ∃con: Connectivity | con ∈ connections
  • (con . connection . 1 . id, con . connection . 2 . id)
  = (route i, route (i + 1)) ∧ (∀cono: Connectivity | cono ∈ connections
  • con . weight ≤ cono . weight)
∀ alter: seq Node | alter ∈ alternatives ∧ # alter > 1
  • ∀i: ℕ | i ∈ 1 .. # alter - 1
  • ∃con: Connectivity | con ∈ connections
  • (con . connection . 1 . id, con . connection . 2 . id)
  = (alter i, alter (i + 1))
  ∧ (∀cono: Connectivity | cono ∈ connections
  • con . weight ≤ cono . weight)
```

Invariants: (i) The total number of routes in primary and alternating tables are same. It might be possible that there does not exist any alternate route for a particular primary route. In such cases, we have supposed that if an alternate route does not exist then it is defined as an empty sequence and hence the property is satisfied. (ii) The first elements of both the primary and alternate routes are same. Similarly, the last elements are also same. (iii) The sets of internal elements except first and last elements of primary and alternate routes are disjoint. (iv) All consecutive elements in a primary route are connected in the ad hoc network. Further, every node when selected for to be part of the primary route is the most appropriate for transmission of data. (v) All consecutive elements in an alternate route are connected in the graph relation and every node of the alternate route is the most appropriate for transmitting the data.

After description of the routing table its management is required if a node moves from one location to another and is a part of any communication between two nodes, or a new route is established. Only one operation to add a primary and alternate route is defined here and other operations can be defined similarly. A route is searched only if it does not exist in the routing table. After route is established and data can be transmitted, the route is stored in the routing table for its future use. After discovery of primary and alternate routes the routing table is updated by the schema ARRT given below. The schema takes routing table and new established routes as input and updates the routing table as an output.

ARRT

ΔSPRTable

ΔAPRTable

route?, alter?: seq Node

```
route? ∉ routes
alter? ∉ alternatives
# route? > 1 ∧ # alter? > 1
route? 1 = alter? 1 ∧ route? (# route?) = alter? (# alter?)
∀i: ℕ | # route? > 1 ∧ i ∈ 2 .. # route? - 1
  • ∀j: ℕ | # alter? > 1 ∧ j ∈ 2 .. # alter? - 1 • route? i ≠ alter? j
∀i: ℕ | # alter? > 1 ∧ i ∈ 2 .. # alter? - 1
  • ∀j: ℕ | # route? > 1 ∧ j ∈ 2 .. # route? - 1 • alter? i ≠ route? j
∀node: Node | node ∈ ran route?
  • ∃object: Object | object ∈ objects • object . id = node
∃object1, object2: Object | object1 ∈ objects ∧ object2 ∈
  objects • # route? ≥ 1 ⇒ route? 1 = object1 . id
  ∧ route? (# route?) = object2 . id
  ∧ object1 . type = Source ∧ object2 . type = Destination
∀node: Node | node ∈ ran alter?
  • ∃object: Object | object ∈ objects • object . id = node
∃object1, object2: Object | object1 ∈ objects ∧ object2 ∈
  objects • # alter? ≥ 1 ⇒ alter? 1 = object1 . id
  ∧ alter? (# alter?) = object2 . id
  ∧ object1 . type = Source ∧ object2 . type = Destination
routes' = routes ∪ {route?}
alternatives' = alternatives ∪ {alter?}
```

Invariants: (i) The newly established route is not already in the routing table. (ii) The newly established alternate route does not exist in the alternate routing table. (iii) Both primary and alternate routes have at least two nodes. (iv) Primary and alternate route are disjoint. (v) For every node in the route there is an object with same identity. (vi) For every route there exists two objects, one is source and the other is destination. (vii) For every node in the alternate route there is an object with same identity. (viii) For every alternate route there exists two objects, one is source and the other is destination. (ix) The new state of primary routing table is union of its previous routes and the newly established route. (x) The new state of alternate routing table is union of its previous alternate routes and the newly established route.

VI.CONCLUSION

In this paper, a formal procedure of defining and managing routing table for mobile Ad Hoc network is presented. For every route an alternate route is supposed if it exists. Initially, routing table is represented by a graph relation and then transformed to Z notation. An object of the network is represented by a node and communication link is assumed as an edge of the graph. Because objects are free to move from one place to another consequently the communication links might be active at one time and dead at another time. Hence the network is not fixed and it needs frequent management to update the active nodes and live communication links. For this purpose efficient and optimal procedures are needed to optimize and improve the quality of service.

First of all basic data types, node, communication link, battery status etc. are defined at a highest level of specification. Then moving objects and connectivity are defined to construct the entire network as a graph relation. Based on the definition of network, formal description of ad hoc network is presented. Then routing table is defined for the primary routes and alternate route is assumed for each primary route if exists. Graph theory is used in this research because it has several applications in modeling of communication networks and was proved an effective tool for this problem. On the other hand, it does not have much computer tool support for verifying and validating the computer models. Formal methods are approaches based on mathematical techniques and have a rigorous computer tools support used for analysis, specification and proving of the computerized models. That is why integration of graph theory and formal methods in terms of Z notation is used in this research.

After formalizing the procedures for AODV protocol, it was observed that inconsistencies and ambiguities were identified and removed by application of formal methods for the specification of the routing table. We believe that this integrated approach will be useful for further analysis and optimization of other similar routing protocols used for communication.

REFERENCES

- [1] D. B. Johnson, D. A. Maltz, and J. Broch, "Dynamic source routing for multihop wireless ad hoc networks", Ad Hoc Networking, Addison-Wesley, pp 139-172, 2001.
- [2] C. E. Perkins, "Ad hoc on demand distance vector (AODV) routing", <http://www.ietf.org/internetdrafts/>, IETF Internet Draft, 1998.
- [3] V. Park and S. Corson, "Temporally-ordered routing algorithm (TORA) Version 1 functional specification", IETF Internet Draft, work in progress, draft-ietf-manet-tora-spec-04.txt, 2001.
- [4] S. R. Das, R. Casteneda, J. Yan, "Simulation based performance evaluation of mobile, ad hoc network routing protocols", ACM Mobile Networks and Applications (MONET), 2000.
- [5] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols", proceedings of MobiCom, pp. 85-97, 1998.
- [6] P. Johansson, T. Larson, N. Hedman, B. Mielczarek and M. DegerMark, "Scenario-based Performance Analysis of Routing Protocols for Mobile Ad hoc Networks", in Proceedings of MOBICOM, 1999.
- [7] S. Ahmed and A. K. Ramani "Alternate route for improving quality of service in mobile ad hoc networks", International Journal of Computer Science and Network Security, 11(2), pp. 41-50, 2011
- [8] D. Cavin, Y. Sasson, and A. Schiper. "On the accuracy of MANET simulators", Proceedings of the Workshop on Principles of Mobile Computing, pp. 38-43, 2002.
- [9] J. M. Spivey, "The Z notation: a reference manual", Prentice Hall, 1989.
- [10] X. Niu, Z. Tao, G. Wu, C. Huang, Li Cui, "Hybrid cluster routing: an efficient routing protocol for mobile ad Hoc networks", IEEE International Conference on Communications, 2006.
- [11] L. Wang and S. Olariu, "A two-zone hybrid routing protocol for mobile ad hoc networks", IEEE Transactions on Parallel and Distributed Systems, 15(12), 2004.
- [12] Charles E. Perkins, Elizabeth M. Belding-Royer, and Samir R. Das. "Ad hoc on demand distance vector (AODV) routing", Internet draft, 2003.
- [13] S. R. Das "On-demand multi-path distance vector routing for ad hoc networks", IEEE ICNP, 2001.
- [14] Z. Qiang Ye, S.V. Krishnamurthy and Satish K. "A framework for reliable routing in mobile ad hoc networks", IEEE INFOCOM, 2003.
- [15] S. J. Lee and M. Gerla: "AODV-BR: Backup Routing in Ad Hoc Networks", IEEE WCNC 2000, 2000.
- [16] C. Woo Ahn, S. Hwa Chung, T. Hun Kim, Su-Young Kang, "A node-disjoint multi-path routing protocol based on AODV in mobile ad-hoc networks", International Conference on Information Technology, IEEE, 2010.
- [17] W. K. Lai, S.-Y. Hsiao, and Y.-C. Lin. "Adaptive backup routing for ad-hoc networks", Computing Communication 2, pp. 453-464, 2007.
- [18] A. Hall. "Realising the benefits of formal methods", Journal of Universal Computer Science, 13(5), pp.669-678, 2007.
- [19] S. Chiyangwa and M. Kwiatkowska, "Modeling ad hoc on-demand distance vector (AODV) protocol with time automata", Proceedings of Third Workshop on Automated Verification of Critical Systems, 2003.
- [20] J.M. Spivey, "The Z notation, a reference manual", 2nd edition, Prentice Hall International, 1992.
- [21] J. Woodcock, J. Davis; "Using Z specification, refinement and proof", Prentice Hall, 1996.
- [22] Teyseyre, A. "A 3D visualization approach to validate requirements", Proceedings Congreso Argentino de Ciencias de la Computacion, 2002.
- [23] ISO: Information technology, "Z formal specification notation, syntax, type system and semantics", ISO/IEC International, 2002.

Speed Response and Performance Degradation of High Temperature Gamma Irradiated Silicon PIN Photodiodes

Abd El-Naser A. Mohamed¹, Nabil A. Ayad², Ahmed Nabih Zaki Rashed^{1*} and Hazem M. El-Hageen^{1,2}

¹Electronics and Electrical Communication Engineering Department,
Faculty Electronic Engineering, Menouf, 32951, Egypt

²Atomic Energy Authority, P.O. Box 29, Naser City, Cairo, Egypt

*E-mail: ahmed_733@yahoo.com

Abstract—In the present paper, we have been investigated deeply and parametrically the speed response of Si PIN photodiodes employed in high temperature-irradiated environment. The radiation-induced photodiodes defects can modify the initial doping concentrations, creating generation-recombination centres and introducing trapping of carriers. Additionally, rate of the lattice defects is thermally activated and reduces for increasing irradiation temperature as a result of annealing of the damage. Nonlinear relations are correlated to investigate the current-voltage and capacitance-voltage dependences of the Si PIN photodiodes, where thermal and gamma irradiation effects are considered over the practical ranges of interest. Both the ambient temperature and the irradiation dose possess sever effects on the electro-optical characteristics and consequently the photo-response time and SNR of Si PIN photodiodes. In this paper, we derive the transient response of a Si PIN photodiode for photogeneration currents, when it is exposed to gamma radiation at high temperature. An exact model is obtained, which may be used to optimize the responsivity and speed of these irradiated devices over wide range of the affecting parameters.

Keywords: Radiation effects, PIN photodiode, Optoelectronics, Transient current, Dark current, Photocurrent, Quantum efficiency, Responsivity.

I. INTRODUCTION

Since several years, photonic and lightwave technology is seriously considered for optical access communication and monitoring applications in space borne local communication systems and nuclear projects. A major problem in these environments is the presence of radiation fields. Two types of damage affect the electronic devices when they are exposed to the radiation [1]. The first one is ionization damage, it is a transitory damage. In contrast, displacement damage is considered permanent. For several reasons, the interest of study of the effects of performance of devices in high-temperature electronics is developing rapidly. If these components are to be used in a radiation environment, knowledge about the degradation under high-temperature irradiation conditions is highly desirable [2]. Optical communication devices in close proximity to radiation fields such as those present in terrestrial orbits and high-energy accelerators suffer both long term effects due to total dose and displacement damage from non-ionizing energy-loss, as well as short term or transient effects due to local ionization from nuclear reactions and high-energy recoils generated close to or within the depletion volume of a device. A photodiode works by generating current from photons absorbed in its active area. In a semiconductor material, a region depleted of mobile charge carriers is formed near the P-N junction. This zone is called depletion region. Incident radiation within this region will create electron-hole pairs, immediately separated

by the internal field. With no external voltage applied, this internal field will prevent the majority carriers to cross the junction. Minority carriers however are still capable of reaching the junction by diffusion and give rise to leakage current. Electron-hole pairs generated outside the depletion region will most likely recombine, consequently not contributing to the photocurrent [3].

The current-voltage technique is used to measure the rate of carrier creation and so the generation or the recombination rate [4]. On the other hand, the capacitance-voltage technique in reverse bias direction, is used to determine doping profiles of a semiconductor [4–6]. The capacitance measurements give information about fixed impurity states and defect centers in the band gap. Device testing, adequate system shielding and radiation tolerant design are some fundamental steps in the methodology or in the radiation hardness assurance [7] that are needed to ensure the correct performance and efficiency of electronics during system life. But, there is an increasing interest in the development of accurate modeling and simulation techniques to predict device response under different radiation conditions [8].

In the present study, we have been investigated and analyzed parametrically and numerically the modeling basics of a PIN photodiode device with the maximum possible precision, in order to predict the frequency response behaviour of Si PIN photodiodes when they were irradiated to different dose of high-temperature gamma radiation environments over wide range of the affecting parameters.

II. Physical Modeling Basics

Radiation damage produces defects which can result in three main effects on photodiode devices as following:

- The increase in dark current can be related to the minority carrier lifetime of the semiconductor if the generation-recombination is dominated by mid-band levels caused by defects. Another source affecting the dark current could be ionizing damage to the surface of the device.
- Degraded photocurrent as defects act as electron or hole trapping centers for the photogenerated pairs. The defects may be primary defects, i.e. defects which originate directly from atomic displacements, or secondary defects resulting from the interaction of mobile primary defects with impurities. Many defects will recombine leading to an immediate repair of the lattice. However, some will combine to form stable defects such as interstitials, di-vacancies, vacancy-impurity complexes, vacancy-dopant complexes, and larger clusters. These defects form effective recombination and trapping centers resulting in a decrease in the minority carrier lifetime, carrier density and carrier mobility. Defect centers position in the band gap determines their activity and hence the

conduction mechanism in devices made from such material [9]. Deep traps are defects whose ionization energy, E , is much greater than $k_B T$ (k_B is the Boltzmann constant and T is the temperature). They trap free carriers with the consequence that they reduce the conductivity considerably. In contrast, shallow traps are easily ionized at equilibrium since $\Delta E \ll k_B T$, and so they increase the conductivity by releasing trapped carriers. In depleted regions they contribute to the space charge and the voltage required for full depletion. Generation-recombination (g-r) centers are situated near the centre of the band gap, in which position their trapping for electrons and for holes is comparable, and so they easily generate or recombine e-h pairs. Then the free carriers are removed to reduce the conductivity. Defect centers can also act as compensation centers in the electrical neutral bulk of a semiconductor. Here, the deep levels are not easily ionized at equilibrium and have the effect of locking away free carriers to reduce the conductivity. The response degradation [8] is probably related to type inversion of the low-doped layer from n to p-type. At low integrated fluence, the radiation forming acceptor state levels compensate the donor states until the effective doping concentration N_{eff} is reduced to that of the intrinsic semiconductor. At higher fluences, the effective doping is mainly provided by the radiation induced defects. The concentration of majority carriers decreases with the irradiation fluence.

- Degraded rise and fall times due to de-trapping or a reduction in the carrier mobility [10]: The decrease in photocurrent and the increase in the dark current are expected to be the major changes in thin junction devices such as photodiodes. The change in the device response, rise and fall times are expected to be small, but still require measuring. Three main factors limit the speed of response of a photodiode. These are [11]:

- a) The drift time of the carriers through the depletion region;
- b) The diffusion time of the carriers generated outside the depletion region;
- c) The time constant incurred by the capacitance of photodiode with its load and its associated circuit.

Photons that penetrate the semiconductor can be absorbed and its energy can be utilized in the generation of e-h pairs. The model that describes the rate of generation is [12]:

$$G_{opt}(x) = (1-r)\eta \frac{P_0 \alpha}{h\nu} \exp(-\alpha x) \quad (1)$$

where r is the reflection coefficient, Previous reports in literature have stated that is independent of dose for 1 MeV electron irradiations up to $5 \times 10^{15} \text{ cm}^{-2}$ [13]. η is the quantum efficiency, P_0 is the incident light intensity, h is the Planck constant, ν is the photon frequency, α is the absorption coefficient and x is the depth variable. The optical spectral response of a PIN photodiode is called the optical sensitivity or the responsivity and it is related to the total photon-induced current. If the width of the p-layer is much thinner than $1/\alpha$, the photon-induced current in the p-layer does not contribute to the total photocurrent.

The current-voltage and capacitance-voltage characteristics, in the dark and under illumination are highly sensitive to the radiation-induced change of the minority carrier lifetime τ . In general, the damage coefficients for the mean minority carrier lifetime in semiconductors depend on

the following parameters; type and energy of the incident particle, kind of material, resistivity, types and concentration of impurities, injection level, temperature and elapsed time after irradiation [14, 15]. Also a semiconductor p-n junction acts as a capacitor. The depletion region capacitance of a uniformly doped lifetime diode at full depletion may be expressed in terms of the dielectric constants ϵ_0 , ϵ_r . In this situation the effective carrier concentration is evaluated:

$$N_{eff} = \frac{2C^2}{q\epsilon_r\epsilon_0 A^2} V \quad (2)$$

Where A is the active diode area, q is the electronic charge and V is the full depletion voltage. This relation shows that $N_{eff} \propto VC^2$, which may be simplified to $V \propto C^{-2}$ for a constant effective carrier concentration, which is the case for uniform doping and is assumed for lifetime material. In any semiconductor, a rise in temperature will increase the current, since carriers become thermally activated to increase the effective carrier density, N_{eff} , so that the current $I \propto N_{eff}$. An increase in light intensity is expected to have the same effect [9]. Because the current is ohmic and is generated in the whole of the depletion region, the depletion width becomes a function of depletion voltage. The capacitance becomes a function of radiation and temperature since electrons and holes are thermally activated.

III. Modeling Description

A) Optical and electrical properties analysis

The dark current, I_D , for a device having depletion depth W , active area A and the effective carrier concentration, N_{eff} under high temperature irradiation T and gamma radiation fluence γ is given by [15, 16]:

$$I_D = \frac{qAW(T, \gamma)N_{eff}(T, \gamma)}{2\tau_r(T, \gamma)} \quad (3)$$

Where τ_r is the minority carrier lifetime after irradiation and it is given by [9, 12]:

$$1/\tau_r = 1/\tau_0 + K_r \gamma \quad (4)$$

Where τ_0 denotes the pre-irradiation minority carrier lifetime respectively, and K_r is the damage coefficient for τ_r . Assuming a linear relationship between damage increase and fluence, the damage coefficient for dark current K_D and light photocurrent K_P , can be defined by following equation [15]:

$$\Delta I_{D,P} = |I_{D,P}(\gamma) - I_{D,P}(0)| = K_{D,P} \gamma \quad (5)$$

A simple model of the annealing can be constructed if we assume that the radiation-induced defects anneal according to a first-order mechanism (exponential recovery) [17], at a given absolute high temperature irradiation T , K_D can be related to an activation energy E by the Arrhenius formula:

$$K_D(T) = K_D(0) \exp(E / K_B T) \quad (6)$$

Where K_B , is Boltzmann's constant.

Based on the data of [18-20], the following nonlinear thermal and radiation relations for the set Si PIN photodiode:

$$I_{Dark, Photo}(T, \gamma) = I_{Dark, Photo}(T) \times I_{Dark, Photo}(\gamma) \quad (7)$$

$$\left. \begin{aligned} I_{Dark} &= I_{Dark}(0) + 3.29 \times 10^{-24} \exp\left(\frac{831}{T}\right) \times \gamma \\ I_{Photo} &= I_{Photo}(0) - 7.8 \times 10^{-14} \left\{ \begin{aligned} &-1.137 \times 10^{-11} T^2 \\ &+ 6.536 \times 10^{-9} T \\ &+ 8.207 \times 10^{-8} \end{aligned} \right\} \times \gamma \end{aligned} \right\} \quad (8)$$

The drift current density of PIN photodiode is given as:

$$j_{drift} = q \int_0^W G_{opt}(x) dx = q \phi_0 (1 - e^{-\alpha W}) = \frac{q \phi_0 (1 - r_f)}{ahv} (1 - e^{-\alpha W}) \quad (9)$$

The one-dimension diffusion equation for initial minority holes P_{n0} in the bulk n region is [11, 21]

$$D_p \frac{\partial^2 P_n}{\partial x^2} - \frac{P_n - P_{n0}}{\tau_p} + G_{opt}(x) = 0 \quad (10)$$

$$P_n^2 - \left(\frac{1}{L_p^2}\right) P_n = -\frac{P_{n0}}{L_p^2} - \frac{\phi_0 \alpha}{D_p} e^{-\alpha x} \quad (11)$$

$$D_p = \frac{L_p^2}{\tau_p} = \frac{K_B T}{q} \mu \quad (12)$$

Where τ_p , L_p and D_p are minority carrier life time, diffusion length and diffusion constant and μ is the carrier mobility. The solution of Eq. 11 under the boundary conditions $P_n = P_{n0}$ for $x = \infty$, and $P_n = 0$ for $x = W$.

$$j_{diffu} = q P_{n0} \frac{D_p}{L_p} + q \phi_0 \left(\frac{\alpha L_p}{1 + \alpha L_p} \right) e^{-\alpha W} \quad (13)$$

$$j_{total} = j_{drift} + j_{diffu} \quad (14)$$

$$j_{total} = q \phi_0 \left(1 + \frac{\alpha L_p - 1 - \alpha L_p}{1 + \alpha L_p} e^{-\alpha W} \right) + q P_{n0} \frac{D_p}{L_p} \quad (15)$$

Under normal operating conditions, the dark-current term involving p_{n0} is much smaller so that the total photocurrent is proportional to the incident photon flux per unit area, ϕ_0 . But under irradiation condition the total photocurrent is given by [9]:

$$I_{total} = I_{photo} + I_{Dark} \quad (16)$$

Where I_{total} is the current measured under illumination, I_{Dark} is the current measured in the dark, I_{photo} is the current due to the illumination only. The high-temperature irradiation induce diffusion length change can be expressed as the following expression:

$$I_{photo} = j_{photo} \times a = \frac{aq \phi_0 (1 - r_f)}{hv} \left(1 - \frac{e^{-\alpha W}}{1 + \alpha L_p} \right) \quad (17)$$

$$L_p = \left(\frac{1}{\alpha_s} \right) \frac{1 - e^{-\alpha W} \left(1 - \frac{I_{photo}}{aq \phi_0} \right)}{e^{-\alpha W} \left(1 - \frac{I_{photo}}{aq \phi_0} \right)} \quad (18)$$

Where a is the photodiode area. By comparing the different cases of the depletion layer width, the junction capacitance and the inverse of the absorption coefficient, a reasonable compromise between high-frequency response and high quantum efficiency of photodiode is found for absorption region thicknesses between $1/\alpha$ and $2/\alpha$.

$$\alpha W = \frac{W}{(1/\alpha)} = -\ln \left((1 + \alpha_s L_p) \left(1 - \frac{I_{photo}}{aq \phi_0} \right) \right) \quad (19)$$

Irradiation induced change of the depletion layer width and the absorption coefficient must be take into consideration. Based on Eq.2 and the results of [5, 6] which shows the variation of the effective carrier concentration, N_{eff} of Si PIN photodiode with electron irradiation dose. The depletion layer capacitance with its initial value C_0 , when a voltage V is applied to a junction with the built-in potential ($V_{bi(Si)} \sim 0.65V$ [9]), is given by:

$$C(T, \gamma) = \frac{C_0 \sqrt{\exp(-\beta \gamma)}}{a_1} \times \frac{(1 + a_2 T)}{\sqrt{|V| + V_{bi}}} \quad (20)$$

Where T is the ambient temperature, γ is irradiation fluence, $C_0/a_1 = 1.176 \times 10^{-9}$, $a_2 = 0.001052$, $\beta = 1.139 \times 10^{-15}$ and $V = -1$ volt. The depletion width W can be expressed as the following [9]

$$W = \sqrt{\frac{2 \epsilon (|V| + V_{bi})}{q N_{eff}}} = \sqrt{2 \epsilon \rho \mu (|V| + V_{bi})} \quad (21)$$

$$W = \frac{\epsilon_0 \epsilon_r A}{C} \quad (22)$$

Where A is the effective photodiode area, q is the electronic charge, N_{eff} is the carrier concentration, ϵ_r is the dielectric constant of silicon, ϵ_0 is the vacuum permittivity. From the previous results [4, 5], we can observed that ϵ_r is constant for radiation but it is function of temperature as [22]:

$$\epsilon_r(T)_{Si} \approx 11.631 + 1.0268 \times 10^{-3} T + 1.0384 \times 10^{-6} T^2 - 8.1347 \times 10^{-10} T^3 \quad (23)$$

The Tauc model [23] has been used as a stander model whereby the optical gap of an amorphous semiconductor may be determined as:

$$\alpha(hv) = \alpha_0 (hv - E_g)^2 \quad (24)$$

Where $1/\alpha_0$ is the band edge parameter and E_g is energy gap. The energy gap of the perfect silicon as a function of the temperature is given by [24]:

$$E_g = 1.166 - \frac{4.731 \times 10^{-4} T^2}{636 + T} \quad (25)$$

But for imperfect semiconductor as a result of radiation induced defects the energy band gap E_g is replace by Tauc bandgap energy E_{gTauc} [23], then α will became:

$$\alpha_{Tauc}(hv) = \alpha_0 (hv - E_{gTauc})^2 \quad (26)$$

In this case the residual absorption near the bandgap due to the intragap is called the Urbach tail [25], and can be expressed with the following equation close to the bandgap:

$$\alpha_{Urb}(hv) = A_0 \exp \left(\frac{(hv - E_g)}{E_{Urb}} \right) \quad (27)$$

We need a function for α that is valid for the entire spectral range, i.e. an equation that combines (15) and (16) is smooth at the cross point, E_{cross} :

$$\alpha_{Tauc}(E_{cross}) = \alpha_{Urb}(E_{cross}) \quad (28)$$

$$\alpha'_{Tauc}(E_{cross}) = \alpha'_{Urb}(E_{cross}) \quad (29)$$

Where α' denotes the first derivative with respect to the energy. With equations (28) and (29) the following conditions are obtained:

$$E_{gTauc} = E_g - 2E_{Urb} \left(1 + \ln \left(\frac{1}{2E_{Urb}} \sqrt{\frac{\alpha_0}{A_0}} \right) \right) \quad (30)$$

$$E_{Urb}(T) = E_{u0} \times T \quad (31)$$

Where E_{u0} and Urbach parameters of silicon material are $A_0 = 800 \text{ cm}^{-1}$, $E_{u0} = 36 \text{ MeV}$ [25], and $\alpha_0 = 4685 \text{ cm}^{-1}$ [26].

B) Photodiode response analysis

The responsivity, S , of a PIN photodiode can be expressed as:

$$S(T, \gamma) = \frac{I_{photo}}{P_0} = \frac{q \eta}{hv} \quad (32) \text{ Wh}$$

ere the quantum efficiency, η , can be given by:

$$\eta(T, \gamma) = \frac{I_{photo} / q}{P_0 (1 - r) / hv} = (1 - R_f) \left(1 - \frac{e^{-\alpha W}}{1 + \alpha L_p} \right) \quad (33)$$

In order to analyze the response time of irradiated PIN photodiode, assume a modulated photon flux density as:

$$\varphi = \varphi_0 \exp(j\omega t) \text{ photons } / (s.cm^2) \quad (34)$$

To fall on photodiode, where ω is the sinusoidal modulation frequency. The total current through the depletion region generated by this photon flux can be shown to be [21, 27]:

$$J_{total} = \left(\frac{j\omega\epsilon(|V| + V_{bi})}{W} + q\varphi_0 \frac{1 - e^{-j\omega t_{dr}}}{j\omega t_{dr}} \right) e^{j\omega t} \quad (35)$$

$$\left| \frac{J_{total}}{aq\varphi_0} \right| = \sqrt{\left(\frac{\sin^2\left(\frac{\omega t_{dr}}{2}\right)}{\left(\frac{\omega t_{dr}}{2}\right)^2} \left(1 - \frac{\omega\epsilon(|V| + V_{bi})}{W(\omega t_{dr})^2} \right) + \left(\frac{\omega\epsilon(|V| + V_{bi})}{W} \right)^2 \right)} \quad (36)$$

Where ϵ is the material permittivity, t_{dr} is the transit drift time of carriers through the depletion region is:

$$t_{dr} = W / 2v_d \quad (37)$$

For p on n devices where W is the width of the depletion region and v_d is the average drift velocity of the carriers. In terms of measurable components Eq.37 The transit drift time becomes:

$$t_{dr} = \frac{W^2}{\mu(|V| + V_{bi})} \quad (38)$$

Where μ is the carrier mobility.

The time for diffusion of carriers from the undepleted region to the depleted region is given by:

$$t_{df} = \frac{\ell^2}{2D} \quad (39)$$

Where D and ℓ are the diffusion constant and the undepleted thickness, which changes with the changing of the depletion layer width W . The time constant t_{RC} of the photodiode with a load resistance R_L is given by:

$$t_{RC} = 2.2(R_S + R_L)C \quad (40)$$

Where C is the capacitance of photodiode at applied bias V , R_S is the series resistance of photodiode, it is the resistance of the contacts and the undepleted bulk of the substrate.

$$R_S = \frac{\rho(W_0 - W)}{A} + \text{contact resistore} \quad (41)$$

Where W_0 is the substrate thickness. Finally, for fully depleted photodiodes the rise time t_r and fall time are generally the same.

$$t_r = \sqrt{t_{dr}^2 + t_{df}^2 + t_{RC}^2} \quad (42)$$

Finally the power signal-to-noise ratio (SNR) S/N at the output of an analog optical receiver is defined by [11]:

$$S/N = \frac{0.5m^2 I_{photo}^2}{(2qf_{-dB}(I_{photo} + I_{Dark}) + 4K_B T f_{-dB} F_n / R_L)} \quad (43)$$

Where m is the analog modulation index, $f_{dB} \approx 0.35/t_r$ is bandwidth and t_r is rise time. The term $(4K_B T f_{dB} F_n / R_L)$ is the total noise associated with amplifier, it is referred to thermal noise of load resistor R_L by the amplifier noise figure F_n .

IV. Results and Discussion

Based on the above modeling equations analysis, the dynamic electro-optical characteristics of Si PIN photodiode are processed in high temperature gamma rays irradiations fields. The double impact of thermal and radiation effects are analyzed over ranges of causes (affecting parameters). As $10^{12} > \gamma$, Fluence, $e/cm^2 > 10^{14}$ and $300 > T$, Temperature, $K > 400$. Special software programs are designed, cast and employed to handle the given basic model, where variation of set of electrical and optical devices parameters $\{I_D, I_P, \alpha, \eta\}$ against variations of a set of two effects $\{T, \gamma\}$ are processed. The device parameters are computed on bases of results of [5, 6, 18, 19, 20]. These variations will effect on response time t_r , responsivity S and signal to noise ratio SNR of the device.

Based on the assumes set of the operating parameters, and the equations analysis, then the obtained results are displays in Figs. (1-7), for the processed Si device of optical wavelength of $\lambda = 950$ nm are assured the following facts:

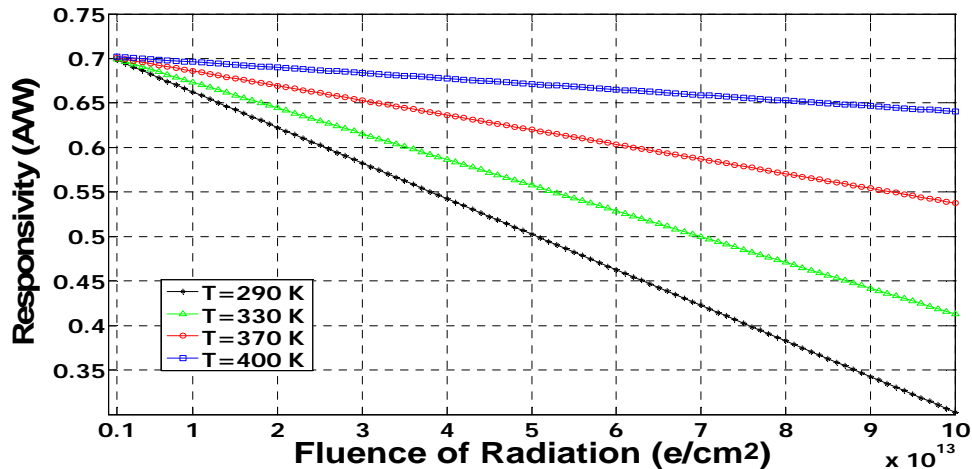


Fig. 1. Optical sensitivity of irradiated Si PIN photodiode at various radiation fluence and temperature.

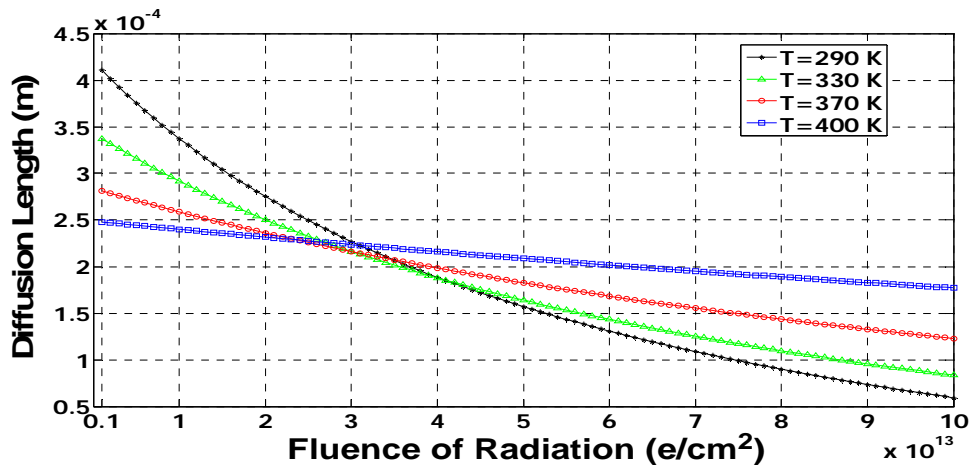


Fig. 2. Diffusion Length of irradiated Si PIN photodiode at various radiation fluence and temperature.

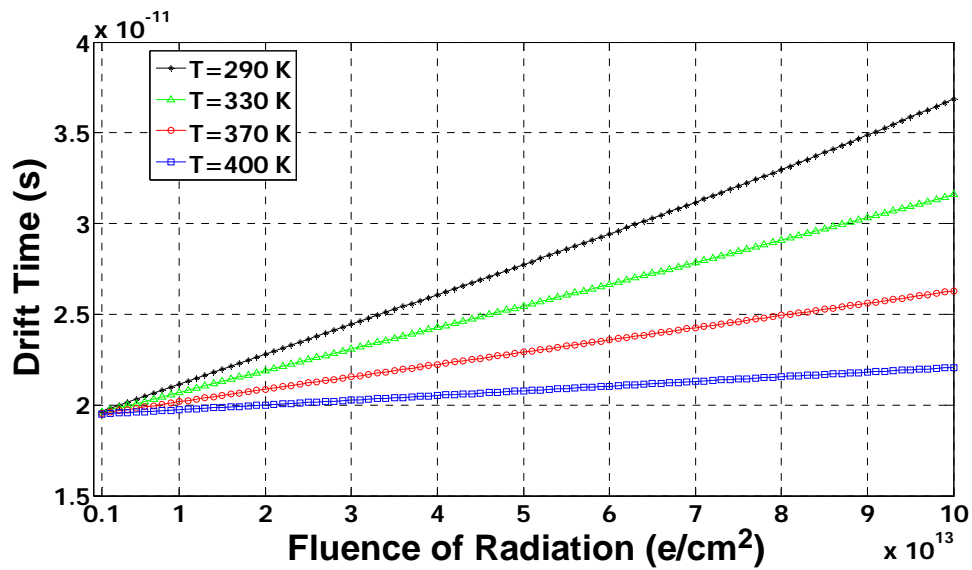


Fig. 3. Drift Time of irradiated Si PIN photodiode at various radiation fluence and temperature.

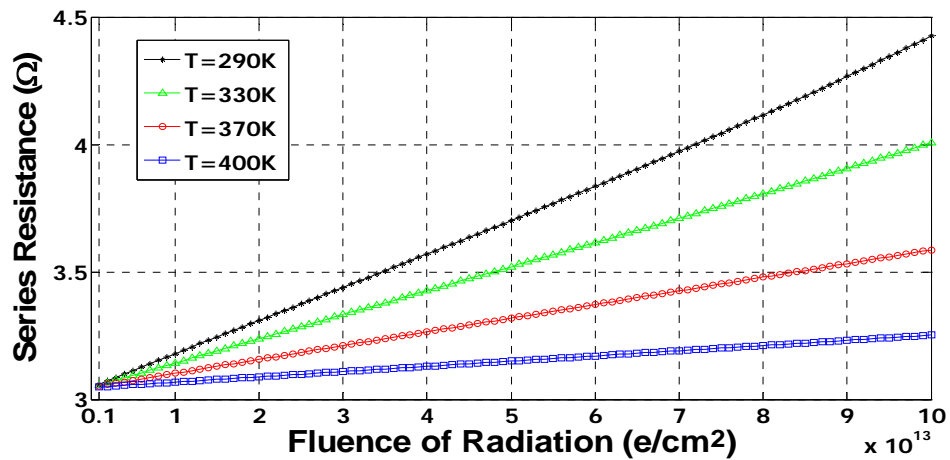


Fig. 4. Series Resistance of irradiated Si PIN photodiode at various radiation fluence and temperature.

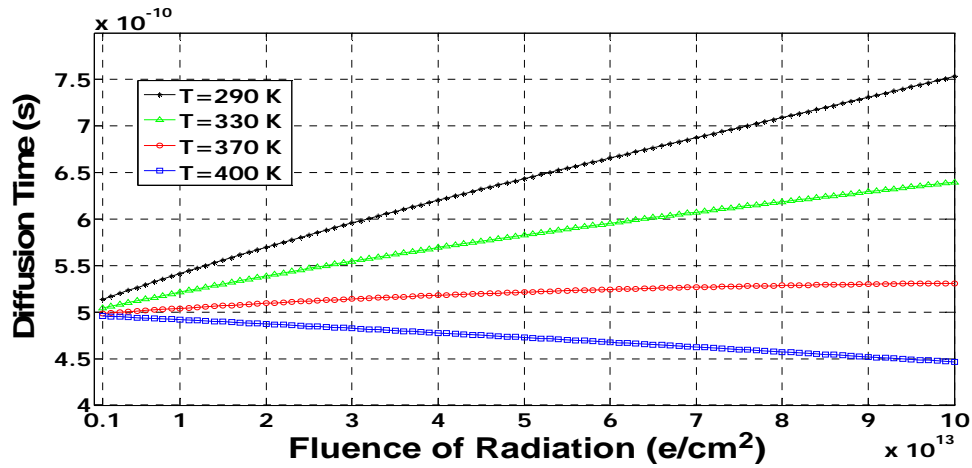


Fig. 5. Diffusion Time of irradiated Si PIN photodiode versus radiation fluence and temperature.

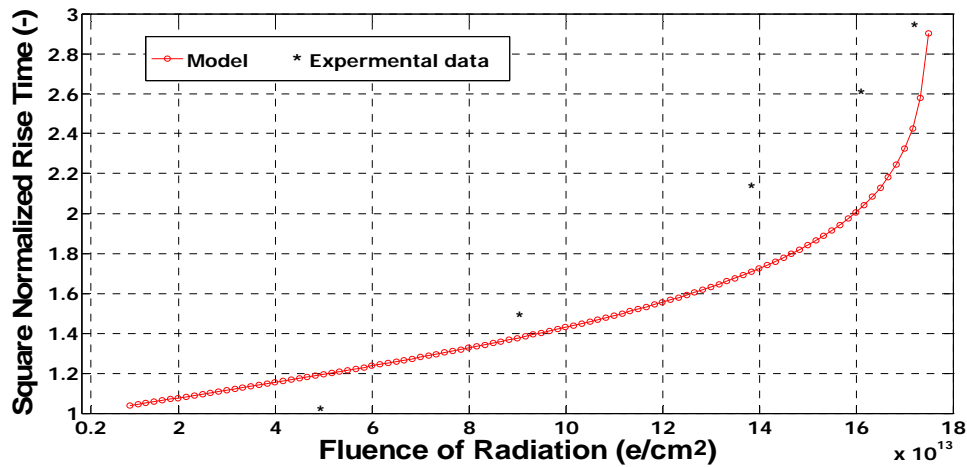


Fig. 6. Square rise time normalized to square rise time of unload Si PIN photodiode as function of radiation fluence.

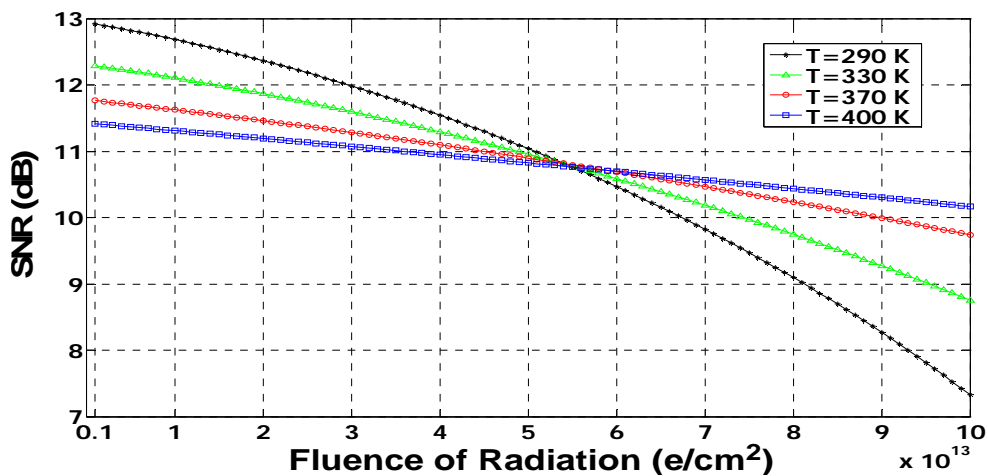


Fig. 7. Plot of SNR of irradiated Si PIN photodiode versus radiation fluence and temperature, $m=0.3$, $F_n=3$ dB, $R_L=50 \Omega$.

- i) As shown in Fig. 1 has assured that as fluence of radiation on optoelectronic Si PIN device increases, this results in responsivity of device decreases at constant ambient temperature. But as ambient temperature increases, this

leads to increase in responsivity of device at constant fluence of radiation.

- ii) Fig. 2 has demonstrated that as fluence of radiation on optoelectronic Si PIN device increases, this results in

diffusion length of carriers decreases at constant ambient temperature. Moreover as ambient temperature increases, this leads to increase in diffusion length of carriers at constant fluence of radiation.

- iii) As shown in Fig. 3 has indicated that as fluence of radiation on optoelectronic Si PIN device increases, this results in drift time of carriers increases at constant ambient temperature. But as ambient temperature increases, this leads to decrease in drift time of carriers at constant fluence of radiation.
- iv) Fig. 4 has proved that as fluence of radiation on optoelectronic Si PIN device increases, this results in series resistance increases at constant ambient temperature. But as ambient temperature increases, this leads to decrease in series resistance at constant fluence of radiation.
- v) As shown in Fig. 5 has indicated that as fluence of radiation on optoelectronic Si PIN device increases, this results in diffusion time of carriers slightly increases at constant ambient temperature. But as ambient temperature increases, this leads to decrease in diffusion time of carriers at constant fluence of radiation.
- vi) Fig. 6 is confirmed that square of normalized rise time follow the empirical data [10] with the our model.
- vii) Fig. 7 has assured that as fluence of radiation on optoelectronic Si PIN device increases, this results in signal to noise ratio decreases at constant ambient temperature. Moreover as ambient temperature increases, this leads to decrease in signal to noise ratio at constant fluence of radiation.

Therefore we can summarized the following results on the SI PIN photodiode performance in the following points:

- a) The damage caused by fluence γ results in decreasing responsivity and signal to noise ratio whatever the set of effect ambient temperature T . But the annealing caused by ambient temperature T , decrease slightly the negative effects of irradiation on optoelectronic device performance.
- b) There is a delay in speed of the response as a result of increasing of the drift and the diffusion times with the increase of the irradiation fluence.
- c) There is an increase in the series resistance R_s of the device with the irradiation fluence.
- d) Radiation defect centers will reduce the minority carrier diffusion length in undepleted region, then the photocurrent will reduce.

V. Conclusions

In a summary, we have presented analytical modeling of the dynamic characteristics of Si PIN photodiode under high temperature gamma radiation. The modeling basics yields an analytical expression for the responsivity of optoelectronic device as a function of irradiation Fluence and temperature. It is evident that also enable better prediction of photocurrent levels, delays and signal bandwidth. Moreover we have demonstrated the circuit effects on signal performance that included as a value of signal to noise ratio. The degradation and delays can be explained by a decrease in the life time and diffusion length of minority carrier caused by the formation of

radiation defects. It is found that the degradation of device performance decrease with increasing irradiation temperature. This result suggests that creation and recovery of the radiation damage proceeds simultaneously at high temperature degrees.

REFERENCES

- [1] Yu. K. Akimov, "Silicon Radiation Detectors: Instruments and Experimental Techniques," IEEE Trans. on Nucl. Sci., Vol. 50, No. 1, pp. 1–28, 2007.
- [2] J.R. Srouf, C.J. Marshall and P.W. Marshall, "Review of Displacement Damage Effects in Silicon Devices," IEEE Trans. on Nucl. Sci., Vol. 50, No. 3, pp. 653–670, 2003.
- [3] A.M. Saad, "Effect of cobalt 60 and 1 MeV electron irradiation on silicon photodiodes/solar cells," J. Phys. Vol. 80, pp.1591–1599, 2002.
- [4] M. Pattabi, S. Krishnan, and G. Sanjeev, "Studies on the temperature dependence of I–V and C–V characteristics of electron irradiated silicon photo-detectors," J. Solar Energy Materials & Solar Cells 91, pp.1521–1524, 2007.
- [5] S. Krishnan, G. Sanjeev, and M. Pattabi, "8 MeV electron irradiation effects in silicon photo-detectors," Nucl. Instr. and Meth. in Physics Research B 264, pp. 79–82, 2007.
- [6] A. Simon, G. Kalinka, M. Jaks, Z. Pastuovic, M. Nova and A.Z. Kiss, "Investigation of radiation damage in a Si PIN photodiode for particle detector," Nucl. Instr. and Meth. in Physics Research B 260, pp. 304–308, 2007.
- [7] M. Van Uffelen, Ivan Genchev and F. Berghmans, "Reliability study of photodiodes for their potential use in future fusion reactor environments," SPIE Proceeding 5465, pp. 92–102, 2004.
- [8] K. Gill, S. Dris, R. Grabit, R. Macias, E. Noah, J. Troska and F. Vasey, "Radiation hardness assurance and reliability testing of InGaAs photodiodes for optical control links in the CMS experiment," IEEE Trans. On Nucl. Sci., Vol. 5, No. 2, pp. 1480–1487, 2005.
- [9] M. McPherson, "Infrared photoconduction in radiation-damaged silicon diodes," J. Opt. A: Pure Appl. Opt. 7, pp. S325–S330, 2005.
- [10] Shinobu Onoda, Toshio Hirao, Hisayoshi Itoh and Tsuyoshi Okamoto, "Evaluation of Transient Current in Si PIN photodiode Induced by High-Energy Charged Particles," Proc. Sch. Eng. Sch. Inf. Sci. Tokai Univ. 31, pp. 1–4, 2006.
- [11] G. Keiser, Optical Fiber Communications, 4th Ed., Tata McGraw-Hill, New Delhi, 2008.
- [12] M.A. Cappelletti, U. Urcola and E.L. Peltzer y Blanc'a., "Radiation-damaged simulation PIN photodiodes," Semicond. Sci. Technol., Vol. 21, pp. 346–351, 2006.
- [13] S. Onoda, T. Hirao, J.S. Laird, H. Mori, T. Okamoto, Y. Koizumi, and H. Itoh, "Spectral Response of a Gamma and Electron Irradiated Pin Photodiode," IEEE Trans. on Nucl. Sci., Vol. 49, No. 3, pp. 1446–1449, 2002.
- [14] B. Danilchenko, A. Budnyk, L. Shpinar, D. Poplavskyy, S.E. Zelensky, K.W.J. Barnham and N.J. Ekins, "1 MeV electron irradiation influence on GaAs solar cell performance," J. Solar Energy Materials and Solar Cells 92, pp. 1336–1340, 2008.
- [15] H. Ohyama, K. Takakura, K. Hayama, S. Kuboyama, Y. Deguchi, and S. Matsuda, E. Simoen and C. Claeys,

- "Damage coefficient in high-temperature particle and γ -irradiated silicon p-i-n diodes," J. Appl. Phys. Lett., Vol. 82, No. 2, pp. 296-298, 2003.
- [16] M. A. Cappelletti, A.P. C'edola and E.L. Peltzer y. Blanc'a, "Simulation of silicon PIN photodiodes for use in space-radiation environments," Semicond. Sci. Technol., Vol. 23, No. 2, pp. 7-13, 2008.
- [17] J.R. Srour, Fellow and D.H. Lo, "Universal Damage Factor for Radiation-Induced Dark Current in Silicon Devices," IEEE Trans. on Nucl. Sci., Vol. 47, No. 6, pp. 2451-2459, 2000.
- [18] K. Takakura, K. Hayama, D. Watanabe, H. Ohyama, T. Kudou, K. Shigaki, S. Matsuda, S. Kuboyama, T. Kishikawa, J. Uemura, E. Simoen and C. Claeys, "Radiation defects and degradation of Si photodiodes irradiated by neutrons at low temperature," Physica B 376-377, pp. 403-406, 2006.
- [19] H. Ohyama, T. Hirao, E. Simoen, C. Claeys, S. Onoda, Y. Takami and H. Itoh, "Impact of lattice defects on the performance degradation of Si photodiodes by high-temperature gamma and electron Irradiation," Physica B 308-310, pp. 1226-1229, 2001.
- [20] H. Ohyama, E. Simoen, C. Claeys, K. Takakura, H. Matsuoka, T. Jono, J. Uemura and T. Kishikawa, "Radiation damage in Si photodiodes by high-temperature Irradiation," Physica E 16, pp. 533 - 538, 2003.
- [21] S. M. Sze, Physics of Semiconductor Devices, John Wiley & Sons, 3rd Ed, USA, 2007.
- [22] H. Li, "Refractive Index of Silicon and Germanium and Its Wavelength and Temperature Derivatives, Journal of Physical and Chemical Reference Data," Vol. 9, No. 3, pp. 561-658, 1980.
- [23] J. Tauc, Amorphous and liquid semiconductors, New York, Plenum Press, 1974.
- [24] Benjawan Kjomrattanawanich, Raj Korde, Craig N. Boyer, Glenn E. Holland, and John F. Seely, "Temperature Dependence of the EUV Responsivity of Silicon Photodiode Detectors," IEEE Trans. on Electron Devices, Vol. 53, No. 2, pp.218-223, 2006.
- [25] Y. Pan, F. Inam, M. Zhang, and D. A. Drabold, "Atomistic Origin of Urbach Tails in Amorphous Silicon," J. Physical Review Letters, Vol. 20, No. 6, pp. 1-4, 2008.
- [26] Martin A. Green, "Self-consistent optical parameters of intrinsic silicon at 300 K including temperature coefficients," J. Solar Energy Materials & Solar Cells 92, pp. 1305- 1310, 2008.
- [27] K. Konno, O. Matsushima, D. Navarro, and M. Miura-Mattausch, "High frequency response of p-i-n photodiodes analyzed by an analytical model in Fourier space," J. Appl. Phys., Vol. 96, No. 7, pp. 3839-3844, 2004.

Author's profile



Dr. Ahmed Nabih Zaki Rashed

was born in Menouf city, Menoufia State, Egypt country in 23 July, 1976. Received the B.Sc., M.Sc., and Ph.D. scientific degrees in the Electronics and Electrical Communications Engineering Department from Faculty of Electronic Engineering, Menoufia University in 1999, 2005, and 2010 respectively. Currently, his job carrier is a scientific lecturer in Electronics and Electrical Communications Engineering Department, Faculty of Electronic Engineering, Menoufia university, Menouf, postal Menouf city code: 32951, EGYPT.

His scientific master science thesis has focused on polymer fibers in optical access communication systems. Moreover his scientific Ph. D. thesis has focused on recent applications in linear or nonlinear passive or active in optical networks. His interesting research mainly focuses on transmission capacity, a data rate product and long transmission distances of passive and active optical communication networks, wireless communication, radio over fiber communication systems, and optical network security and management. He has published many high scientific research papers in high quality and technical international journals in the field of advanced communication systems, optoelectronic devices, and passive optical access communication networks. His areas of interest and experience in optical communication systems, advanced optical communication networks, wireless optical access networks, analog communication systems, optical filters and Sensors, digital communication systems, optoelectronics devices, and advanced material science, network management systems, multimedia data base, network security, encryption and optical access computing systems. He is a reviewer member in high quality scientific research international journals in the field of Electronics, Electrical communication and advanced optical communication systems and communication access networks. His personal electronic mail (E-mail:ahmed_733@yahoo.com).

A Multistage Detection and Elimination of Spurious Singular Points in Degraded Fingerprints

Zia Saquib, Santosh Kumar Soni, Sweta Suhasaria
Center for Development of Advanced Computing
Mumbai, Maharashtra 400049, India
saquib@cdacmumbai.in
santoshsoni@cdacmumbai.in
swetas@cdacmumbai.in

Dimple Parekh, Rekha Vig
NMIMS University,
Mumbai, Maharashtra 400056, India
dimple.parekh@nmims.edu
rekha.vig@nmims.edu

Abstract—Singular point (SP) detection is one of the most crucial phases in fingerprint authentication systems and is used for fingerprint classification, alignment and matching. This paper presents a multistage approach for detection and elimination of spurious singular points especially in degraded fingerprints. The approach comprises three stages. In the first stage, two different methods, viz., quadrant change and orientation reliability measure, are independently employed on the same image to generate two sets of candidate singular points. The second stage performs the multiscale analysis on a set of candidate SPs located by reliability method, which improves the approximation by reducing the list of SPs. In the third stage, the spurious singular points are detected and thereby eliminated by taking the intersection of the two sets of SPs. This model is tested on a proprietary (Lumidigm Venus V100 OEM Module sensor) fingerprint database at 500 ppi resolution. The experimental results show that the approach effectively eliminates the spurious SPs from the noisy and highly translated/rotated fingerprint images. The proposed scheme is also compared with one of the state-of-the-art techniques, the experimental results prove its superiority over the later.

Keywords- Spurious Singular Points, Multiscale Analysis, Orientation Consistency, Quadrant Change, Reliability, Minimum Inertia, Maximum Inertia.

I. INTRODUCTION

The performance of fingerprint authentication system has come a long way but it is still influenced by many factors, like: inaccurate detection of singular points (core and delta). Poor-quality and noisy fingerprint images mostly result in false or missing singular points (SPs), which generally results in degradation of the overall performance of the authentication systems. This paper presents a three-stage approach, which primarily focuses on the detection and elimination of spurious SPs for all types of fingerprint images, especially noisy images. This paper puts forward an effective way to locate a unique reference point consistently and accurately using tri-method fusion scheme. Method-A works on the quadrant change information, whereas, Method-B uses pixel-wise reliability measure of the orientation field followed by multiscale analysis to compute candidate SPs. Intersection of methods A and B

gives the genuine set of SPs. These methods, the proposed scheme and its comparison with one of the state-of-the-art techniques are explained in detail in section II. Experimental results are discussed in sections III, followed by conclusion in section IV.

II. THE PROPOSED SCHEME AND ITS KEY COMPONENTS

A. Quadrant Change: Method-A

As per K. Kryszczuk and A. Drygajlo (2006)[2], a singular point is the location where the general ridge orientation becomes discontinuous. Informally, this can be stated as the area where ridges oriented rightwards change to leftwards and those that were oriented upwards turn downwards, and opposite. This information can be extracted from the quadrant change of the averaged square gradients. The orthogonal gradient components in the x and y directions are considered separately. In general, each pair of corresponding gradient components manifests the gradient quadrant change by the change of sign. The sign maps PM_x and PM_y are computed using the Eq. (1):

$$\begin{bmatrix} PM_x \\ PM_y \end{bmatrix} = \begin{bmatrix} \wedge_{(\overline{G_x}) \neq 0} PM_x = \text{sign}(\overline{G_x}) \\ \wedge_{(\overline{G_x}) = 0} PM_x = 1 \\ \wedge_{(\overline{G_y}) \neq 0} PM_y = \text{sign}(\overline{G_y}) \\ \wedge_{(\overline{G_y}) = 0} PM_y = 1 \end{bmatrix}. \quad (1)$$

We need to locate points in whose respective local ridge gradients change sign in both x and y directions. These points are obtained by computing the intersection of the two sets of such points for which the sign of the y-directional and x-directional (respectively) gradient component changes, as shown in Eq. (2):

$$[x_{sp}, y_{sp}] = \langle x, y : \text{edge}(PM_x) \rangle \cap \langle x, y : \text{edge}(PM_y) \rangle \quad (2)$$

The operator *edge* in Eq. (2) denotes any edge detector that works on binary images, and $[x_{sp}, y_{sp}]$ are the points where two quadrants change boundaries intersect, as shown in Figure 1. $[x_{sp}, y_{sp}]$ are considered as SPs, as shown in Figure 2. This method works well with good quality gray-level images, but the moment image quality degrades, it starts resulting in spurious SPs and eventually becomes ineffective, as shown in Figure 2.

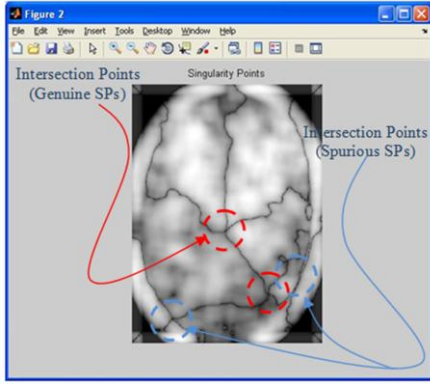


Figure 1. Horizontal and Vertical maps.

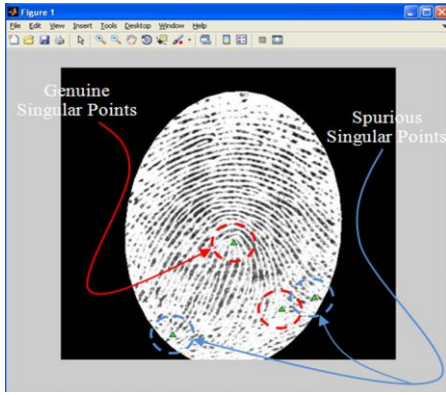


Figure 2. Genuine and Spurious SPs based on Quadrant Change Information.

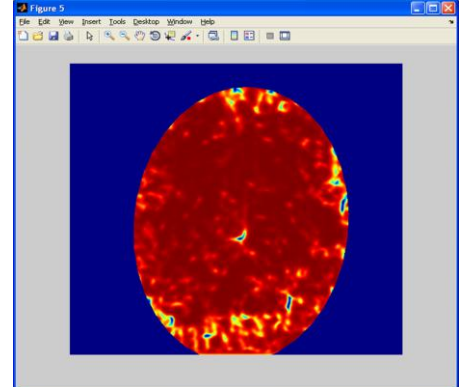


Figure 3. Reliability Image

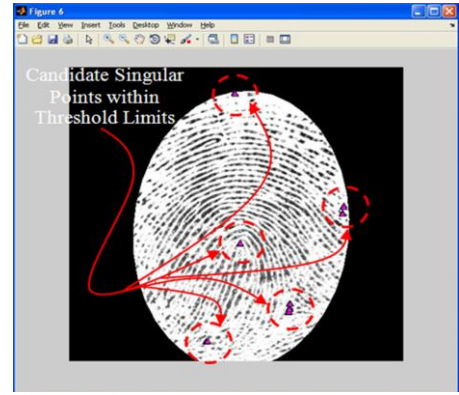


Figure 4. Genuine and Spurious SPs based on Reliability Measure.

B. Orientation Reliability Measure: Method B

As per Z. Saquib and S. K. Soni (2011)[6], M. Khalil, D. Muhammad (2010)[5], the raw fingerprint image is first filtered using Gabor filter. Then, 'reliability' of ridge orientation map is calculated, followed by the calculation of the area of moment of inertia about the orientation axis (the min. inertia) and an axis perpendicular (the max. inertia), as given in Eq. (3) and (4):

$$\min_inertia(x, y) = (((Gyy + Gxx) - (Gxx - Gyy) * \phi'_x) - (Gxy * \phi'_y))/2 \quad (3)$$

$$\max_inertia(x, y) = Gyy + Gxx - \min_inertia(x, y) \quad (4)$$

where, ϕ'_x and ϕ'_y are cosine and sine of doubled angles (ridge orientations). The reliability measure is given by Eq. (5):

$$\text{Reliability Measure} = 1.0 - \min_inertia/\max_inertia \quad (5)$$

All such pixels with reliability measure below an empirically determined threshold (here, it is 0.035) are considered as the candidate SPs. The pixels with deep blue shades are the possible SPs, as shown in Figure 3, and the corresponding SPs are shown in Figure 4, which is inclusive of both genuine and spurious.

C. MultiScale Analysis

As per T. Van and H. Lee (2009)[1], a multiscale analysis (see Figure 5) of orientation consistency is used to search the local minimum orientation consistency from large scale to fine scale. The orientation consistency-based technique can be summarized as follows:

- 1) Compute the orientation consistency $\text{Cons}(s)$ of each block based on the outside 8s surrounding blocks of its $(2s+1) \times (2s+1)$ neighborhood.
- 2) Find the minimum orientation consistency denoted as $\text{Cons}_{\min}(s)$. Compute candidate threshold as,
$$T = \begin{cases} \text{Cons}_{\min}(s) + 0.15 & \text{if } \text{Cons}_{\min}(s) > 0.5 \\ \text{Cons}_{\min}(s) + 0.05 & \text{Otherwise,} \end{cases} \quad (6)$$
- 3) Select the blocks if their $\text{Cons}(s) < T$.
- 4) Compute $dx(s)$ and $dy(s)$, and select the blocks with both $dx(s)$ and $dy(s)$ larger than 0 as the candidate blocks in the next finer scale:

$$dx(s) = \sum_{k=-s}^s \cos(2\theta(i-s, j+k)) - \sum_{k=-s}^s \cos(2\theta(i+s, j+k)), \quad (7)$$

$$dy(s) = \sum_{k=-s}^s \sin(2\theta(i+k, j-s)) - \sum_{k=-s}^s \sin(2\theta(i+k, j+s)). \quad (8)$$

- 5) If no candidate blocks for the reference point are located, let $T = T + 0.01$, go to step 3).
- 6) Repeat steps 1), 2), 3), 4), and 5) in the selected candidate blocks with $s = s-1$ until $s = 1$.
- 7) Locate the block with minimum orientation consistency $\text{Cons}(1)$ from the selected finest scale blocks as the unique reference point.

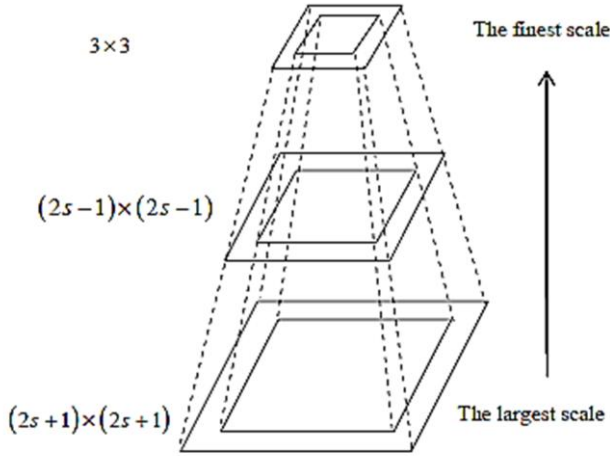


Figure 5. The multiscale analysis of orientation consistency.

We have performed multiscale analysis over the set of SPs given by reliability measure stage for better approximation of the genuine SPs, as explained in sub-section D. Multiscale analysis helps in reducing the list of SPs further by isolating and removing the false SPs.

D. Proposed Approach: A Multistage Detection and Elimination of Spurious SPs

The proposed approach, as shown in Figure 6, comprises the state-of-the-art methods (with some modifications/tuning) presented in sub-sections A, B and C. Firstly, the two sets of candidate SPs are generated using the methods: i) quadrant change information and ii) reliability measure of the orientation field. In order to have better approximation, multiscale analysis is performed over the candidate SPs from reliability measure, which reduces (or minimizes) the list by identifying, and thereby ignoring most of such pixels which are not likely to be the SPs. Finally, the genuine SPs are confirmed by taking the intersection of the two sets of SPs from the above two methods, which then filters out the false SPs, if any, leaving behind genuine SPs. These stages are shown together in Figure 6. The experimental results are shown in Figure 7 and 8. In Figure 8, first column depicts the raw images, second column shows the results using Quality Change and Reliability methods, third column displays SPs by Quadrant Change Information (blue), Reliability Measure (red), Multiscale Analysis (green) and the fourth column presents results from the proposed scheme (genuine SPs are depicted by orange color). Few improved

cases are also presented in Figure 9, where the raw images chosen are relatively of much poorer quality than the images in Figure 8.

III. EXPERIMENTAL RESULTS

Proprietary (Lumidigm Venus V100 OEM Module sensor) dataset has been chosen as test data to evaluate the impact of the proposed multistage scheme for detection and elimination of spurious SPs. The scheme is implemented in MATLAB. The experimental results show that this approach satisfactorily improves the accuracy of detection of correct singular points in noisy and highly transformed (translated/rotated) fingerprint images. Only select cases (highly degraded/translated/rotated) have been chosen to measure the effectiveness of the approach. Few of them are presented in Figure 7 and 8. Some improved cases are also displayed, as shown in Figure 9, where severely distorted/poorly overlapped fingerprint images are chosen, which present real challenges in the fields.

IV. CONCLUSION

Genuine SPs are very crucial towards attaining high accuracy and performance of the authentication systems. Thus, spurious SPs need to be completely removed. In this paper, a multistage scheme is proposed for detection and elimination of spurious singular points, especially in highly degraded, translated and rotated fingerprint images. Experimental results clearly show that the three methods in combination effectively remove (or minimize) the spurious singular points. The scheme is tested against some select difficult cases. Also, this method (fourth column in Figure 8), upon comparison with the approach presented by Z. Saquib, S. K. Soni (2011) (second column in Figure 8), is found better.

ACKNOWLEDGMENT

We wish to extend our sincere thanks to the Department of Information Technology (DIT), Ministry of Communications and Information Technology, Govt. of India, for assigning us a biometric project: "BharatiyaAFIS". This work is carried out as a part of the same project.

REFERENCES

- [1] T. Van and H. Lee, "An efficient algorithm for fingerprint reference-point detection", IEEE 2009.
- [2] K. Kryszczuk and A. Drygajlo, "Singular point detection in fingerprints using quadrant change information", The 18th International Conference on Pattern Recognition (ICPR'06), 2006.
- [3] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition. New York: Springer, 2003.
- [4] L. Hong, Y. Wan, and A. Jain, "Fingerprint image enhancement: algorithm and performance evaluation", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 20, No. 8, 1998.
- [5] M. Khalil, D. Muhammad, M. Khan, Mohammed, "Singular points detection using fingerprint orientation field reliability", International Journal of Physical Sciences Vol. 5(4), pp. 352-357, 2010.
- [6] Z. Saquib, S. Soni, S. Suhasaria, D. Parekh, R. Vig, "A fault-tolerant approach for detection of singular points in noisy fingerprint images", International Journal of Computer Security Issues, Volume 8, 2011.
- [7] http://en.wikipedia.org/wiki/Euclidean_distance
- [8] Kovese PD (2008). MATLAB and Octave Functions for Computer Vision and Image Processing, in School of Computer Science and Software Engineering, The University of Western Australia. Available from <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>.

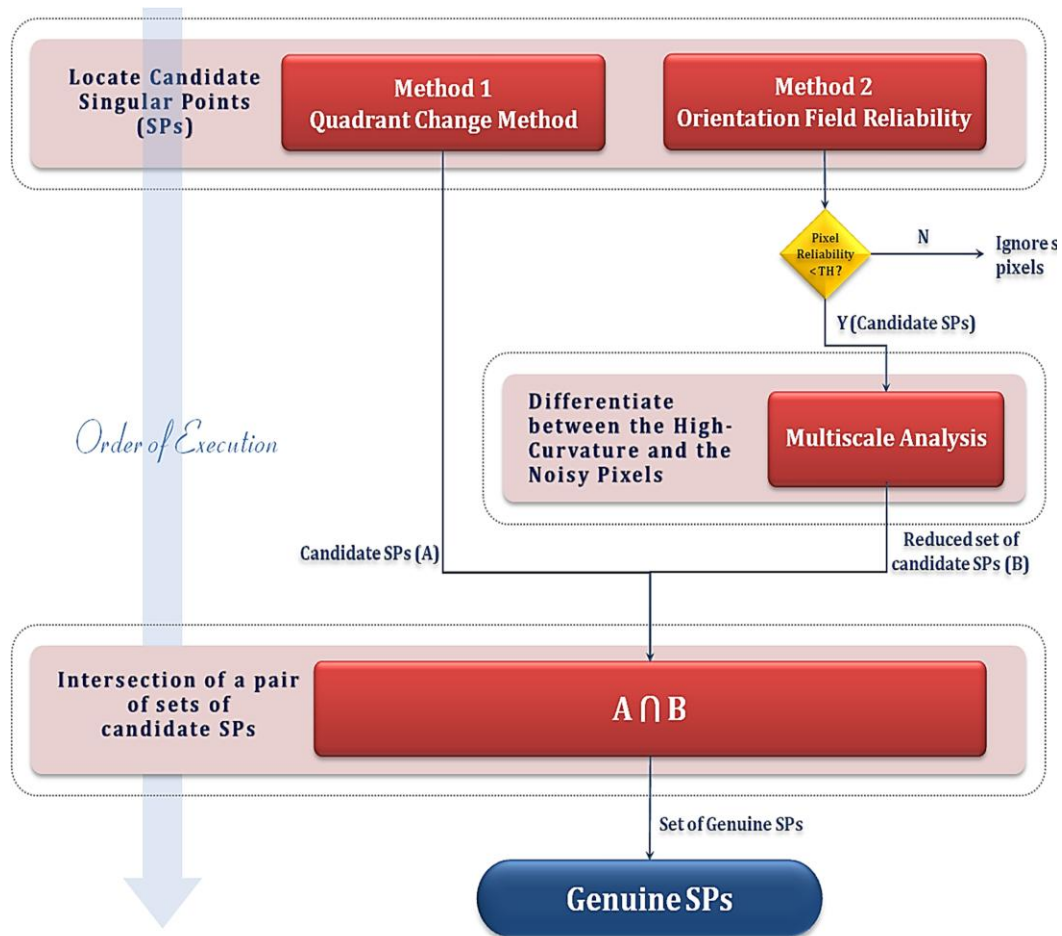


Figure 6. Proposed Scheme.

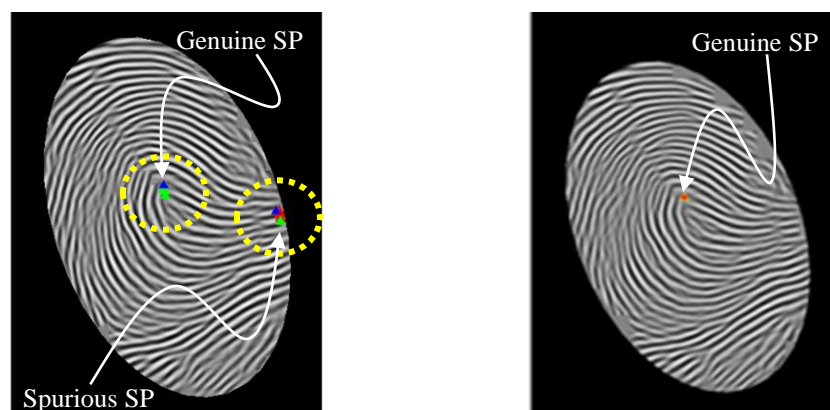
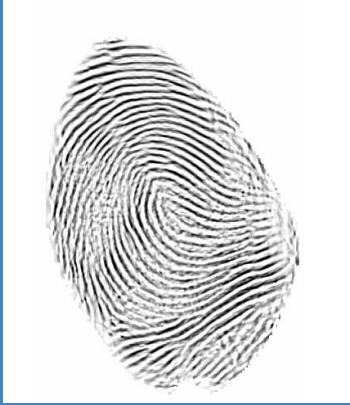

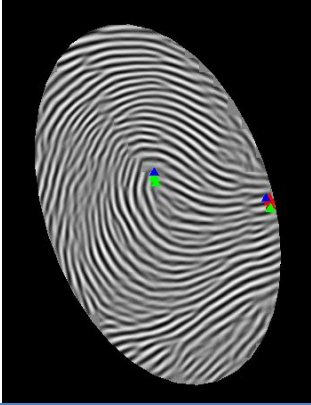

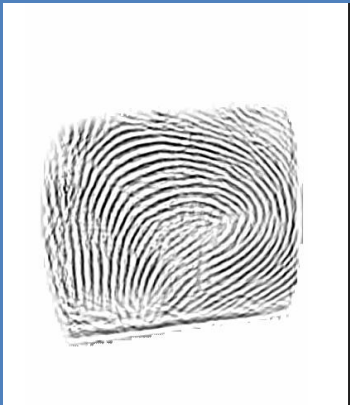
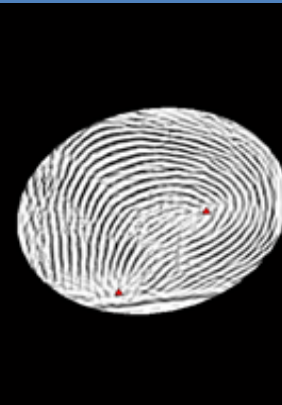
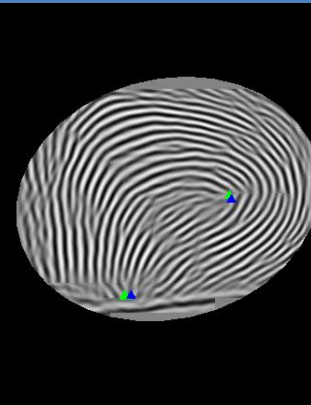
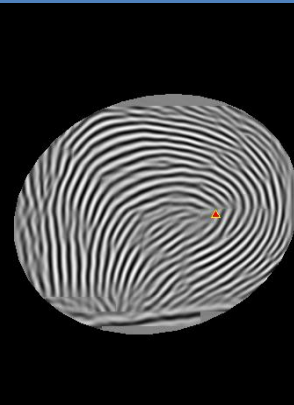


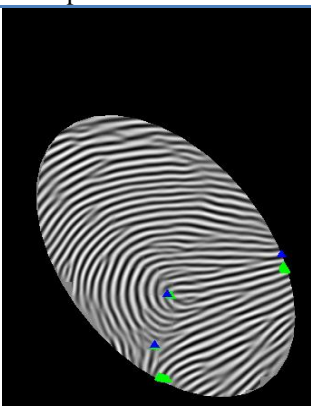
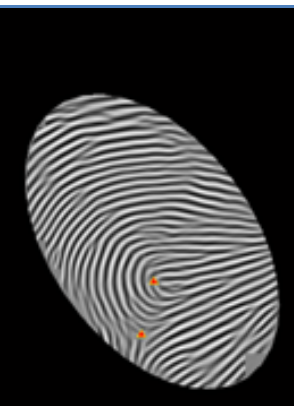
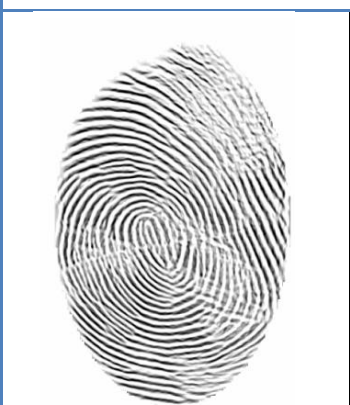

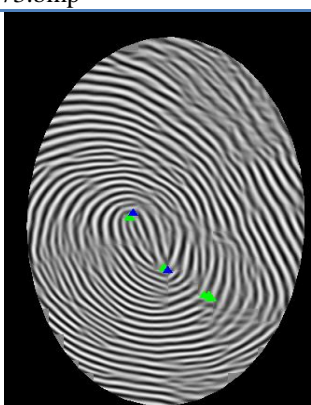

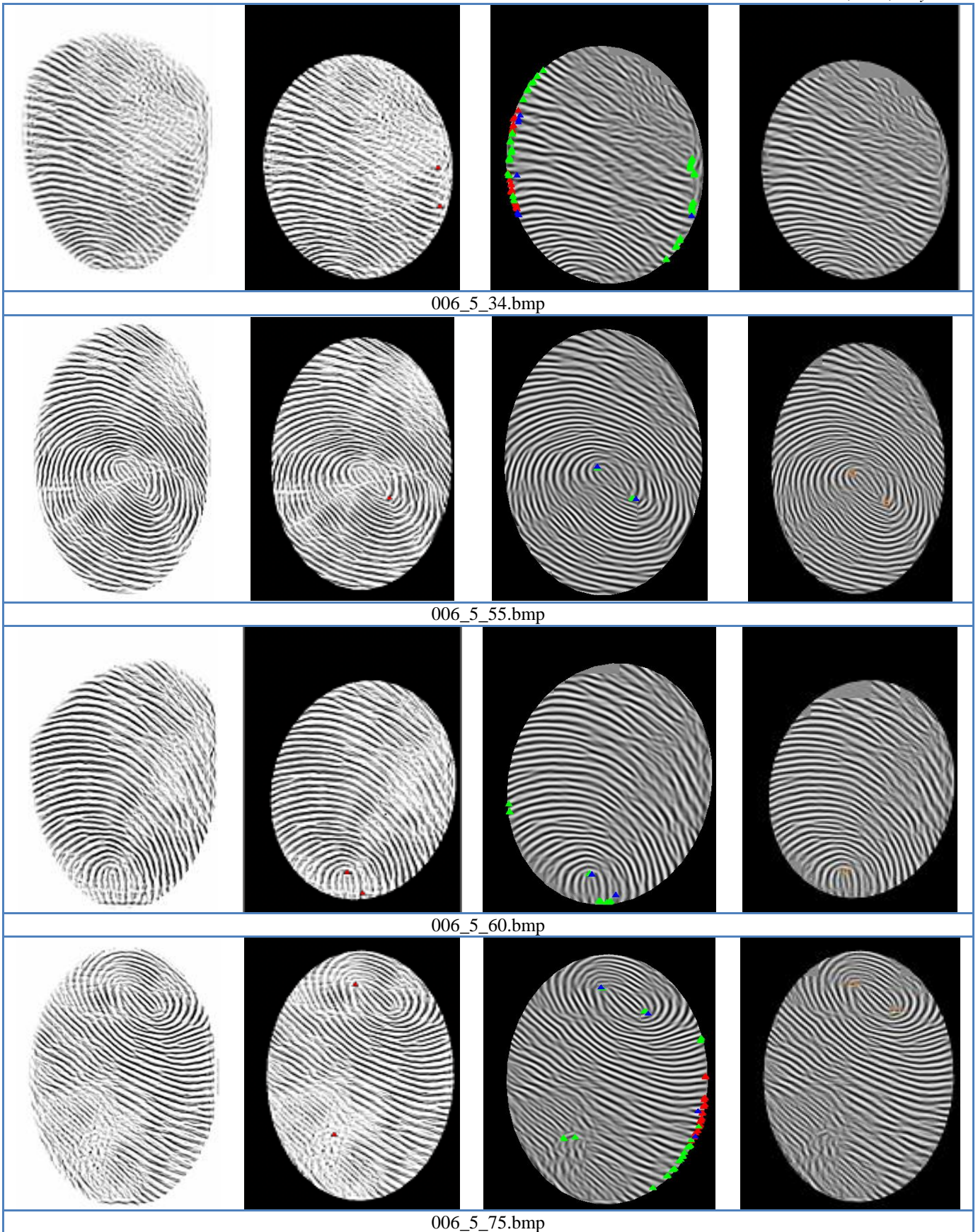


Figure 7. SPs before Intersection (left), SPs after Intersection (right).

Fingerprint Image	Quadrant Change & Reliability methods	Quadrant Change, Reliability & Multiscale methods (before Intersection)	Quadrant Change, Reliability & Multiscale methods (after Intersection)
			
001_5_10.bmp			
			
001_5_68.bmp			
			
003_5_73.bmp			
			
006_5_16.bmp			



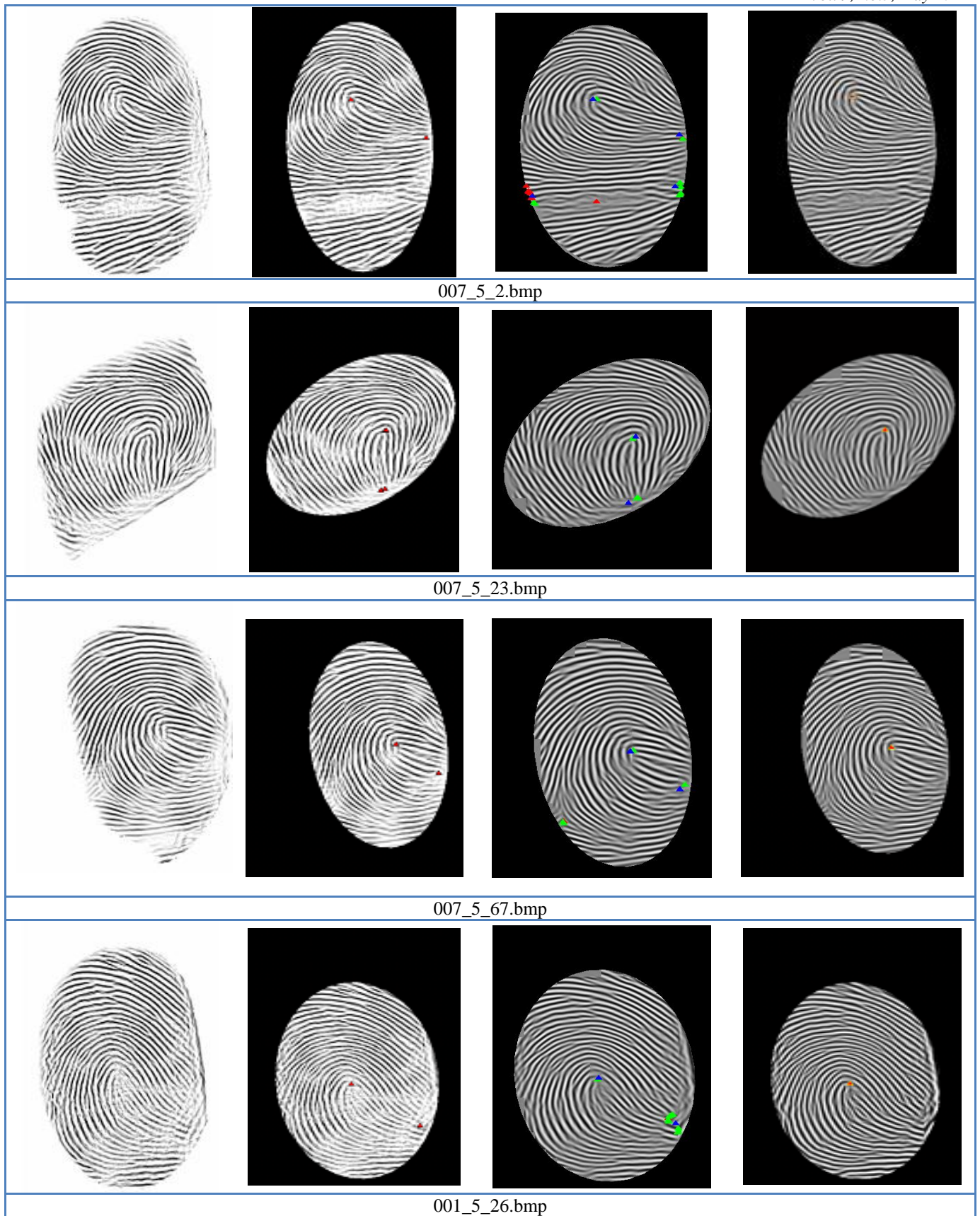


Figure 8. Experimental Results from **Lumidigm Dataset**: (first column) Raw Images, (second column) Results using Quality Change and Reliability methods, (third column) Blue SPs by Quadrant Change Information, Red SPs by Reliability Measure, Green SPs by Multiscale Analysis and (fourth column) Proposed Scheme – Genuine SPs are depicted by Orange SPs.


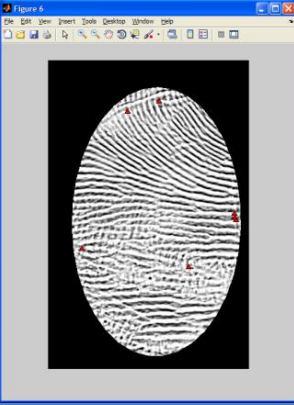
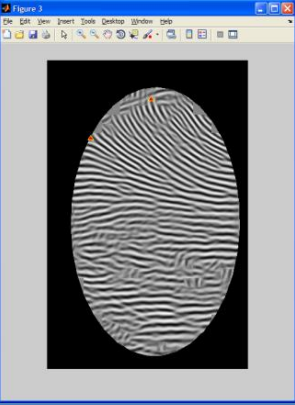


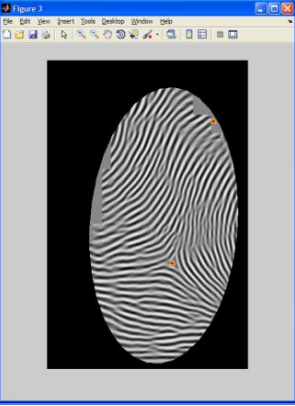

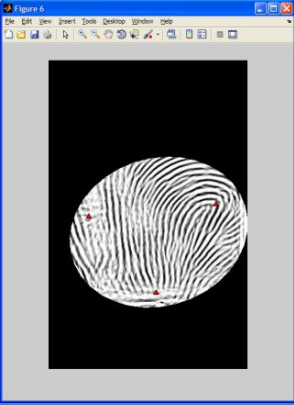




Sr.No.	Fingerprint Image	Quadrant Change & Reliability methods	Quadrant Change, Reliability & Multiscale methods (proposed approach)
1.			
006_5_65.bmp (There is no SP present in the Raw Image)			
2.			
006_5_66.bmp (Only Delta should have been marked)			
3.			
007_5_25.bmp (Only single Core is present)			
4.			
001_5_15.bmp (Only single Core is present)			

Figure 9. Experimental Results from **Lumidigm Dataset**: Third column represent **improved cases**, inclusive of both genuine and spurious SPs (please zoom to view them properly).

Practical Implementation Of Matlab Based Approach For Face Detection Using Feedforward Network

Meenakshi Sharma¹
Sri Sai College Of Engg. & Tech., Pathankot¹
HOD CSE¹
Mss.s.c.e.t@gmail.com

Sukhvinder Singh²
Sri Sai College Of Engg. & Tech., Pathankot² Mtech CSE
4th sem²
sukhaish@gmail.com

Dr. N Suresh Rao³
Jammu University³
HOD MCA³

Abstract- *The objective is to recognise and identify faces, not previously presented to or in some way processed by the system. There are many datasets involved in this project. Some of them are the ORL, MIT database which consisting of a large set of images of different people. The database has many variations in pose, scale, facial expression and details. Some of the images are used for training the system and some for testing. The test set is not involved in any part of training or configuration of the system, except for the weighted committees as described in a section later on.*

Keywords- Face recognition, PCA, Symbols, Matlab, Feedforward Network.

Introduction

The purpose of face Detection algorithm is to examine a set of images and try to find the exact match of a given image. An advanced system would be a neural network face Detection algorithm. The system examines small windows of the image in order to calculate the distances of given points. That would be done from any algorithm but in a system where we use neural networks the system arbitrates between multiple networks in order to improve performance over a single network.

The goal of this ongoing project is to formulate paradigms for detection and Detection[1] of human faces, and especially develop an algorithm, which is going to have high performance in complex backgrounds. One of the applications would be towards adding face-oriented queries to our image database project.

The fundamental principle, which we are exploiting for our face Detection algorithm, is Principal Component Analysis. Though the algorithm is much simpler. One of the aims is to run tests in order to compare the algorithm with two PCA algorithm and also show that the calculation between two given point with the ARENA algorithm is efficient as much as

if we had use the Euclidean distance. The algorithm used for the face Detection, from the project is known as ARENA. Similar to several other approaches to face Detection and identification, which use Principal Component Analysis (PCA) as pre-processing, dimensionality reduction and feature extraction, of the input images. One of the main parts of the project is a neural network. The use of a neural network makes the algorithm perform better.

In chapters two and three we are going to analyse the background of the project. A literature background of face Detection and neural networks discussing also the methods that were used for the project. In the next chapter, four there is a description of the datasets that where used in order to test and train the algorithm and the neural network, the implementation of which is in chapter six. After that, in chapter [2] seven there is a detailed analysis of the outputs we get from the programs and a comparison of the ARENA algorithm with other methods that have been used for face Detection, the theory of which is analysed in chapter five. Finally in chapter eight there is a discussion about the work that had been done and further improvement that could be done.

1. Face Detection:

Face Detection is a part of a wide area of pattern Detection technology. Detection and especially face Detection covers a range of activities from many walks of life. Face Detection is something that humans are particularly good at and science and technology have brought many similar tasks to us. Face Detection in general and the Detection of moving people in natural scenes in particular, require a set of visual tasks to be performed robustly. That process includes mainly three-task acquisition, normalisation and Detection. By the term acquisition we mean the detection and tracking of face-like image patches in a dynamic scene. Normalisation is the segmentation, alignment and normalisation of the face

images[3], and finally Detection that is the representation and modelling of face images as identities, and the association of novel face images with known models.

2. Neural network:

A neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing element or nodes. Neural network architecture is inspired by the architecture of biological nervous systems, which use many simple processing elements operating in parallel to obtain high computation rates .

Neural networks are a form of microprocessor computer system with simple processing elements, a high degree of interconnection, simple scalar messages and adaptive interaction between elements .

The neural networks resemble the brain mainly in two respects

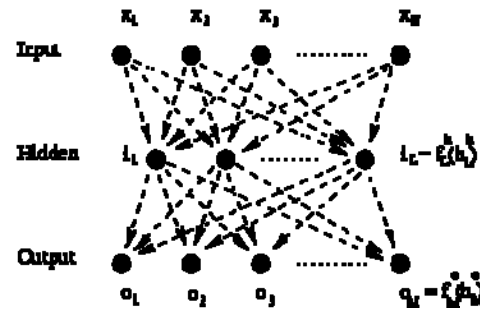
- Knowledge is acquired by the network through a learning process
- Interneuron connection strengths known as synaptic weights are used to store the knowledge.

That means to construct a machine that is able to think. Somehow, not really known yet, the brain is capable to think and perform some operations and computations, much faster sometimes from a computer even the “memory” is much less. How the brain is managed to do that is a hardware parallelism. The computing elements are arranged so that very many of them are working on a problem at the same time. Since there is a huge number of neurones, somehow the weak computing powers of these many slow elements are combined together to form a powerful result.

A Neural Network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns.

HOW ARTIFICIAL NEURAL NETWORKS WORK.

Artificial neural networks can modify their behaviour in the response to their environment. This factor, more than any other, is responsible for the interest they[4] have received. Shown a set of inputs, which perhaps have specific desired output, they self adjust to produce consistent responses. A wide variety of training algorithms has been developed for that reason. Each of the algorithms has it's own strength and weaknesses.

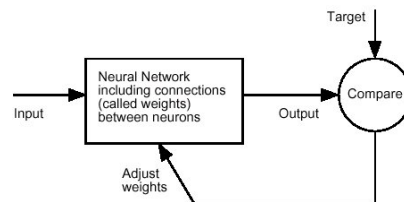


How neural network operate

A. Feedforward Network

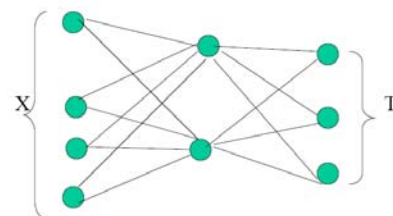
Feedforward networks often have[5] one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with non-linear transfer functions allow the network to learn non-linear and linear relationships between input and output vectors. The linear output layer lets the network produce values outside the range -1 to $+1$. On the other hand, if it is desirable to constrain the outputs of a network then the output layer should use a sigmoid transfer function such as logsig.

Furthermore in the case of multiple-layer networks we use the number of the layers to determine the superscript on the weight matrices. The appropriate notation is used in the two-



layer tansig or purelin network shown next.

Feedforward neural network.



Neural Network

A feed forward network can be used as a general function approximation. It can approximate any function with a finite

number of discontinuities, arbitrarily well, given sufficient neurons in the hidden layer.

In order to create a feedforward neuron network we have to follow a specific procedure. The first step in training a feedforward network is to create the network object. Then we have to initialise the weights and the bias. Then the network is ready to be trained. A feedforward network takes as said, before an object as input and returns[7] a network object with all weights and biases initialised. There is a more detailed analysis of the network and the process we follow in the coding part of the report.

Database Design

The databases, which are used for the project, are standard databases of the University of Surrey, Olivetti-Oracle Research Lab and FERET. Thought is possible to test the algorithm with other databases as well.

The databases consist of more than 400 images, each. All the databases contain images of different people, but in sets. That means that there are a number of images of the same people in each of them. Though each image is different from each other. [6]For example, in the ORL database we have ten different images of each of 40 distinct subjects. For some people, the images were taken at different times, with different lighting, where we might have facial expressions, with open or closed eyes, where the people are smiling or not and facial details, glasses or with out no glasses. Many images of a person can be acquired in a few seconds. Given sufficient data, it becomes possible to model class-conditional structure, i.e. to estimate probability densities for each person.

Apart from that, in all the databases images were taken against a dark homogeneous background with the subjects in an upright, frontal position, but also we have images with more complex backgrounds. On of the most important aspects of the databases is the variation of the pose. There is a limitation of $\pm 20^\circ$ at the posing angle. If the person's pose, in the image is more than then is nearly impossible to be detected from mainly any of the existing face Detection algorithms. Thought in the databases we have posing angle variations but with in the limits.

The files of the images that are used are in TIFF format, and can conveniently be viewed on UNIX, TM systems using the xv program. Most of the images have size of 92x112 pixels, with 256 grey levels per pixel.

II. TEST AND TRAIN SETS

The data sets that have been used for the particular project are divided to two sub sets. The first is the training set that contains the images that were used in order to train the algorithm and the neural network. Training sets are used from the two training programs, arntrn and nntrn. Samples of the set

can be found in appendix II. The other set is the other subset is the testing database, which contains different images than the training set but of the same people. [8]

In MATLAB we use the commands imread and imresize in order to read the images and reduce the resolution. More detailed description of the commands and their properties is given in the code implementation chapter.

3. Algorithms for face Detection

As mentioned in the introduction but also in other parts of the report, there are many algorithms that can be used for face Detection. Most of them are based on the same techniques and methods. Some of the most popular are Principal component analysis and the use of eigenfaces.

A. Principal Component Analysis

On the field of face Detection most of the common methods employ Principal Component Analysis. Principal Component Analysis is based on the Karhunen-Loeve (K-L), or Hostelling Transform, which is the optimal linear method for[9] reducing redundancy, in the least mean squared reconstruction error sense. 1. PCA became popular for face Detection with the success of eigenfaces.

The idea of principal component analysis is based on the identification of linear transformation of the co-ordinates of a system. "The three axes of the new co-ordinate system coincide with the directions of the three largest spreads of the point distributions."

In the new co-ordinate system that we have now the data is uncorrected with the data we had in the first co-ordinate system. [2]

For face Detection, given dataset of N training images, we create N d-dimensional vectors, where each pixel is a unique dimension. The principal components of this set of vectors is computed in order to obtain a d x m projection matrix, W. The image of the i^{th} vector may be represented as weights:

$$\vec{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{im})^T \quad (1)$$

Such that

$$\vec{x}_i = \vec{\mu} + W\vec{\theta} \quad (2)$$

Approximates the original image where μ is the mean, of the x_i and the reconstruction is perfect when $m = d$. P1

As mentioned before the ARENA algorithm is going to be tested and its performance is going to be compared with other

algorithms. For the comparison we are going to use two different PCA algorithms. The first algorithm[11] is computing and storing the weight of vectors for each person's image in the training set, so the actual training data is not necessary. In the second algorithm each weight of each image is stored individually, is a memory-based algorithm. For that we need more storing space but the performance is better.

In order to implement the Principal component analysis in MATLAB we simply have to use the command `prepca`. The syntax of the command is

```
ptrans,transMat = prepca(P,min_frac)
```

`Prepca` pre-processes the network input training set by applying a principal component analysis. This analysis transforms the input data so that the elements of the input vector set will be uncorrected. In addition, the size of the input vectors may be reduced by retaining[10] only those components, which contribute more than a specified fraction (`min_frac`) of the total variation in the data set.

`Prepca` takes these inputs the matrix of centred input (column) vectors, the minimum fraction variance component to keep and as result returns the transformed data set and the transformation matrix.

1) Algorithm

Principal component analysis uses singular value decomposition to compute the principal components. A matrix whose rows consist of the eigenvectors of the input covariance matrix multiplies the input vectors. This produces transformed input vectors whose components are uncorrected and ordered according to the magnitude of their variance.

Those components, which contribute only a small amount to the total variance in the data set, are eliminated. It is assumed that the input data set has already been normalised so that it has a zero mean.

In our test we are going to use two different "versions" of PCA. In the first one the centroid of the weight vectors for each person's images in the training set is computed and stored. On the other hand in PCA-2 a memory based variant of PCA, each of the weight vectors is individually computed and stored.

B. Eigenfaces

Human face Detection is a very difficult and practical problem in the field of pattern Detection. On the foundation of the analysis of the present methods on human face Detection, [12] a new technique of image feature extraction is presented. And combined with the artificial neural network, a new method on human face Detection is brought up. By extraction the sample pattern's algebraic feature, the human face image's

eigenvalues, the neural network classifier is trained for Detection. The Kohonen network we adopted can adaptively modify its bottom up weights in the course of learning. Experimental results show that this method not only utilises the feature aspect of eigenvalues but also has the learning ability of neural network. It has better discriminate ability compared with the nearest classifier. The method this paper focused on has wide application area. The adaptive neural network classifier can be used in other tasks of pattern Detection.

In order to calculate the eigenfaces and eigenvalues in MATLAB we have to use the command `eig`. The syntax of the command is

```
d = eig(A)
```

```
V,D = eig(A)
```

```
V,D = eig(A,'nobalance')
```

```
d = eig(A,B)
```

```
V,D = eig(A,B)
```

`d = eig(A)` returns a vector of the eigenvalues of matrix A . `V,D = eig(A)` produces matrices of eigenvalues (D) and eigenvectors (V) of [13] matrix A , so that $A*V = V*D$. Matrix D is the canonical form of A , a diagonal matrix with A 's eigenvalues on the main diagonal. Matrix V is the modal matrix, its columns are the eigenvectors of A . The eigenvectors are scaled so that the norm of each is 1.0. Then we use `W,D = eig(A')`; $W = W'$ in order to compute the left eigenvectors, which satisfy $W*A = D*W$.

`V,D = eig(A,'nobalance')` finds eigenvalues and eigenvectors without a preliminary balancing step. Ordinarily, balancing improves the conditioning of the input matrix, enabling more accurate computation of the eigenvectors and eigenvalues. However, if a matrix contains small elements that are really due to round-off error, balancing may scale them up to make them as significant as the other elements of the original matrix, leading to incorrect eigenvectors. We can use the no balance option in this event.

`d = eig(A,B)` returns a vector containing the generalised eigenvalues, if A and B are square matrices. `V,D = eig(A,B)` produces a diagonal matrix D of generalised eigenvalues and a full matrix V whose columns are the corresponding eigenvectors so that $A*V = B*V*D$. The eigenvectors are scaled so that the norm of each is 1.0.

C. Euclidean distance

One of the ideas on which face Detection is based is the distance measures, between to points. The problem of finding the distance between two or more point of a set is defined as the Euclidean distance. The Euclidean distance is usually referred to the closest distance between two or more points. So we can define the Euclidean distance d_{ij} between points x_{ik} and x_{jk} as :

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (3)$$

4. Implementation:

The first component of our system is a filter that receives as input a 20x20 pixel region of the image, and generates an output ranging from 1 to -1, signifying the presence or absence of a face, respectively. To detect faces anywhere in the input, the filter is applied at every location in the image. To detect faces larger than the window size, the input image is repeatedly reduced in size (by subsampling), and the filter is applied at each size. This filter must have some invariance to position and scale. The amount of invariance determines the number of scales and positions at which it must be applied. For the work presented here, we apply the filter at every pixel position in the image, and scale the image down by a factor of 1.2 for each step in the pyramid. The filtering algorithm is shown in . First, a preprocessing step, adapted from , is applied to a window of the image. The window is then passed through a neural network, which decides whether the window contains a face. The preprocessing first attempts to equalize the intensity values in across the window. We fit a function which varies linearly across the window to the intensity values in an oval region inside the window. Pixels outside the oval may represent the background, so those intensity values are ignored in computing the lighting variation across the face. The linear function will approximate the overall brightness of each part of the window, and can be subtracted from the window to compensate for a variety of lighting conditions. Then histogram equalization is performed, which non-linearly maps the intensity values to expand the range of intensities in the window. The histogram is computed for pixels inside an oval region in the window. This compensates for differences in camera input gains, as well as improving contrast in some cases. The preprocessed window is then passed through a neural network. Although the figure shows a single hidden unit for each subregion of the input, these units can be replicated. For the experiments which are described later, we use networks with two and three sets of these hidden units. Similar input connection patterns are commonly used in speech and character recognition tasks .The network has a single, real-valued output, which indicates whether or not the window contains a face. The network has some invariance to position and scale, which results in multiple boxes around some faces. To train the [14]neural network used in stage one

to serve as an accurate filter, a large number of face and nonface images are needed. Nearly 1050 face examples were gathered from face databases at CMU, Harvard2, and from the World Wide Web. The images contained faces of various sizes, orientations, positions, and intensities. The eyes, tip of nose, and corners and center of the mouth of each face were labelled manually. These points were used to normalize each face to the same scale, orientation, and position, as follows:

1. Initialize $_F$, a vector which will be the average positions of each labelled feature over all the faces, with the feature locations in the first face F_1 .
2. The feature coordinates in $_F$ are rotated, translated, and scaled, so that the average locations of the eyes will appear at predetermined locations in a 20x20 pixel window.
3. For each face i , compute the best rotation, translation, and scaling to align the face's features F_i with the average feature locations $_F$. Such transformations can be written as a linear function of their parameters. Thus, we can write a system of linear equations mapping the features from F_i to $_F$.
4. Update $_F$ by averaging the aligned feature locations $F_0 i$ for each face i .
5. Go to step 2.

The alignment algorithm converges within five iterations, yielding for each face a function which maps that face to a 20x20 pixel window. Fifteen face examples are generated for the training set from each original image, by randomly rotating the images (about their center points) up to 10°, scaling between 90% and 110%, translating up to half a pixel, and mirroring. Each 20x20 window in the set is then preprocessed (by applying lighting correction and histogram equalization). A few example images are shown in Fig. 4. The randomization gives the filter invariance to translations of less than a pixel and scalings of 20%. Larger changes in translation and scale are dealt with by applying the filter at every pixel position in an image pyramid, in which the images are scaled by factors of 1.2. Practically any image can serve as a nonface example because the space of nonface images is much larger than the space of face images. However, collecting a “representative” set of nonfaces Rowley, Baluja, and Kanade: Neural Network-Based Face Detection (PAMI, January 1998) 4 is difficult. Instead of collecting the images before training is started, the images are collected during training, in the following manner:

1. Create an initial set of nonface images by generating 1000 random images. Apply the preprocessing steps to each of these images.
2. Train a neural network to produce an output of 1 for the face examples, and -1 for the nonface examples. The training

algorithm is standard error backpropagation with momentum. On the first iteration of this loop, the network's weights are initialized randomly. After the first iteration, we use the weights computed by training in the previous iteration as the starting point.

3. Run the system on an image of scenery *which contains no faces*. Collect subimages in which the network incorrectly identifies a face (an output activation > 0).

4. Select up to 250 of these subimages at random, apply the preprocessing steps, and add them into the training set as negative examples. Go to step 2.

Stage Two: Merging Overlapping Detections and Arbitration

The raw output from a single network will contain a number of false detections. In this section, we present two strategies to improve the reliability of the detector: merging overlapping detections from a single network and arbitrating among multiple networks.

Merging Overlapping Detections

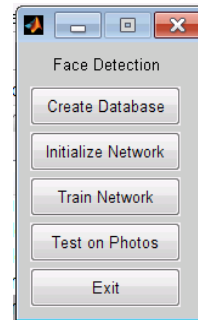
Most faces are detected at multiple nearby positions or scales, while false detections

often occur with less consistency. This observation leads to a heuristic which can eliminate many false detections. For each location and scale, the number of detections within a specified neighborhood of that location can be counted. If the number is above a threshold, then that location is classified as a face. The centroid of the nearby detections defines the location of the detection result, thereby collapsing multiple detections. In the experiments section, this heuristic will be referred to as "thresholding". If a particular location is correctly identified as a face, then all other detection locations which overlap it are likely to be errors, and can therefore be eliminated. Based on the above heuristic regarding nearby detections, we preserve the location with the higher number of detections within Rowley, Baluja, and Kanade: Neural Network-Based Face Detection a small neighborhood, and eliminate locations with fewer detections. In the discussion of the experiments, this heuristic is called "overlap elimination".

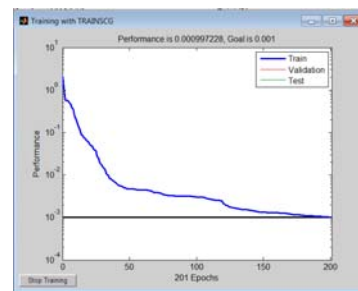
Each detection at a particular location and scale is marked in an image pyramid, labelled the "output" pyramid. Then, each location in the pyramid is replaced by the number of detections in a specified neighborhood of that location. This has the effect of "spreading out" the detections. A threshold is applied to these values, and the centroids (in both position and scale) of all above threshold regions are computed. All detections contributing to a centroid are collapsed down to a single point. Each centroid is then examined in order, starting from the ones which had the highest number of detections within the specified neighborhood. If any other centroid locations represent a face overlapping with the current

centroid, they are removed from the output pyramid. All remaining centroid locations constitute the final detection result. In the face detection work described in [1], similar observations about the nature of the outputs were made, resulting in the development of heuristics similar to those described above.

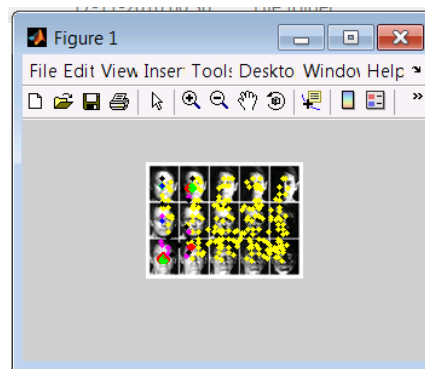
5. Results:



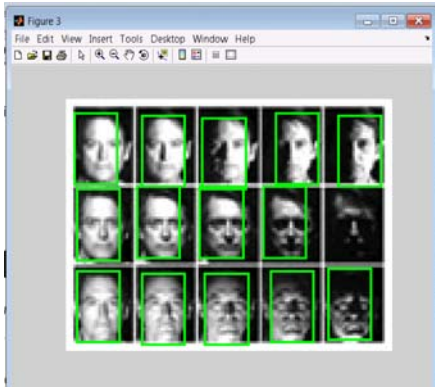
Main Page



Training



Recognising Images



Recognised Faces

6. References

- [1] "Rafael C. Gonzalez and Richard E. Woods", Digital Image Processing, 2nd Edition, , Prentice Hall, 2002.
- [2] Korean standard:
<http://www.cwi.nl/~dik/english/codes/stand.html#ksc>
- [3] "Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression", J.D. Daugman, IEEE Trans. Acoustics, Speech, and Signal Processing, 36:1169-1179, 1988.
- [4] "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex.", J.P. Jones and L.A. Palmer. , J. Neurophysiol., 58(6):1233-1258, 1987.
- [5] "A tutorial on Principal Components Analysis", Lindsay I Smith, February 26, 2002.
- [6] "Face Detection using fisherface algorithm and elastic graph matching", Hyuyng-Ji Lee, Wan-Su Lee, and Jae-Ho Chung, Dept. of Electronic Engr., Inha Univ., Inchon 402-751, Korea
- [7] "A Comparative Study On Color Edge Detection", Andreas Koschan, Proceedings 2nd Asian
- [8] "Chinese Optical Character Detection for Information Extraction from Video Images", Wing Hang Cheung, Ka Fai Pang, Michael R. Lyu, Kam Wing Ng, Irwin King , Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
- [9] "Text Localization, Enhancement and Binarization in Multimedia Documents", ¹Christian Wolf, ¹Jean-Michel Jolion, ²Francoise Chassaing, ¹Lab. Reconnaissance de Formes et Vision, ²France T'el'ecom R&D.
- [10] "Data Mining", Course CSC5180 Web Site , The Chinese University of Hong Kong – The Department of Computer Science and Engineering : <http://www.cse.cuhk.edu.hk/~lwchan/teaching/csc5180.html>
- [11] "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception", Biological Cybernetics, vol. 25, pp. 181.194, 1977.
- [12] "Gray-Scale Character Detection by Gabor Jets Projection", Hiroshi Yoshimura, Minoru Etoh, Kenji Kondo and Naokazu Yokoya, Graduate School of Information Science, Nara Institute of Science and Technology Multimedia Laboratories, NTT DoCoMo, Inc
- [13] "Field. Relations between the statistics of natural images and the response properties of cortical cells", David J. Field, Optical Society of America A, 4(12):2379-2394, 1987.
- [14] "Statistics of Natural Images: Scaling in the wood", Daniel L. Ruderman & William Bialek, Physical Review Letters, 73(6):814-817, 1994.

REDUCING FALSE ALERTS USING INTELLIGENT HYBRID SYSTEMS

Sravan Kumar Jonnalagadda
D.M.S.S.V.H. College of Engineering
Department of Information Technology
Machilipatnam, Andhra Pradesh, INDIA
Email: jnvsravankumar @gmail.com

Subha Sree Mallela
D.M.S.S.V.H. College of Engineering
Department of Computer Science
Machilipatnam, Andhra Pradesh, INDIA

Abstract: Currently, computer systems manage large amounts of data over the network. The growth of data communications has involved an increase in unauthorized accesses and data manipulation with the resulting security violations.. Hackers and intruders have made many successful attempts to bring down high-profile company networks and web services. Many methods have been developed to secure the network infrastructure and communication over the Internet, among them the use of firewalls, encryption, and virtual private networks. Since it is impossible to predict and identify all the vulnerabilities of a network, and penetration into a system by malicious intruders cannot always be prevented, intrusion detection systems (IDS) are essential entities for ensuring the security of a networked system. . An IDS is software (or hardware) designed to detect unwanted attempts at accessing, manipulating, or disabling of computer systems, mainly through a network.. This paper begins with a review of the most well-known intrusion detection technique called snort. The aim of this paper is to present an anomaly detection processor that extends Snort to a hybrid scheme. Finally, the design of a distributed HIDS is proposed that consists of a group of autonomous and cooperating agents. Distributed hybrid intrusion detection system comprising of a multi-agent framework with computational intelligent techniques to reduce the data features to create lightweight detection systems and a hybrid intelligent system approach to improve the detection accuracy.

Keywords- *Network Security, Intrusion detection systems, tcpdump, Snort, Rule structure, Hybrid IDS, anomaly detection processor, Episode Rules...*

I INTRODUCTION

A computer system should provide *confidentiality, integrity* and *assurance* against denial of service. However, due to increased connectivity (especially on the Internet), and the vast spectrum of financial possibilities that are opening up, more and more systems are subject to attack by intruders [1]. These subversion attempts try to exploit flaws in the operating system as well as in application programs and have resulted in spectacular incidents like the Internet Worm incident of 1988.

There are two ways to handle subversion attempts. One way is to prevent subversion itself by building a completely secure system. We could, for example, *require* all users to identify and authenticate themselves; we could protect data by various cryptographic methods and very tight access control mechanisms. However this is not really feasible because:

bugs in popular programs and operating systems that seems to indicate that (a) bug free software is still a dream and (b) no-one seems to want to make the effort to try to develop such software. Apart from the fact that we do not seem to be getting our money's worth when we buy software, there are also security implications when our E-mail software, for example, can be attacked. Designing and implementing a totally secure system is thus an extremely difficult task.

2. Cryptographic methods have their own problems. Passwords can be cracked, users can lose their passwords, and entire crypto-systems can be broken.
3. Even a truly secure system is vulnerable to abuse by insiders who abuse their privileges.
4. It has been seen that that the relationship between the level of access control and user efficiency is an inverse one, which means that the stricter the mechanisms, the lower the efficiency becomes.

The history of security research has taught us a valuable lesson – no matter how many intrusion prevention measures are inserted in a network, there are always some weak links that one could exploit to break in.

We thus see that we are stuck with systems that have vulnerabilities for a while to come. If there are attacks on a system, we would like to detect them as soon as possible (preferably in real-time) and take appropriate action. This is essentially what an Intrusion Detection System (IDS) does. An IDS does not usually take preventive measures when an attack is detected; it is a reactive rather than pro-active agent.

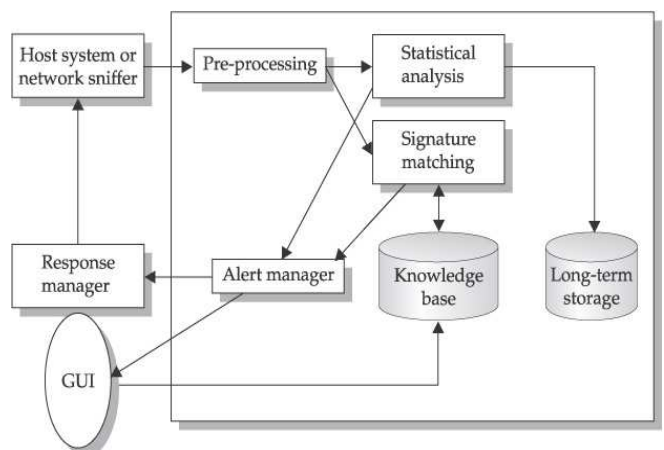


Figure 1. Intrusion Detection System.

1. In practice, it is not possible to build a completely secure system. Miller gives a compelling report on

II BACKGROUND OF INTRUSION DETECTION [2]

Many network problems involve events whose effects are distributed over a wide geographical network area. There are a large number of Intrusion Detection Software / Systems (IDS) out there for various operating platforms, all ranging in price and complexity.

For example, the outbreak of a worm in the Internet can affect traffic patterns across a geographically disparate set of sub networks. The traditional approach to detecting this type of distributed event is to use a set of monitoring systems that are connected to a centralized intrusion detection system. Each monitoring system monitors a separate sub network for any evidence of a distributed event. Any suspicious measurements that are potential evidence for the event of interest are then reported to a centralized intrusion detection system for correlation and further analysis. Hardware and software solutions for a Windows platform found one product that stands out from the rest, is SNORT. SNORT is an open source *Intrusion*.

In the last three years, the networking revolution has finally come of age. More than ever before, we see that the Internet is changing computing, as we know it. The possibilities and opportunities are limitless; unfortunately, so too are the risks and chances of malicious intrusions.

It is very important that the security mechanisms of a system are designed so as to *prevent* unauthorized access to system resources and data. However, completely preventing breaches of security appear, at present, unrealistic. We can, however, try to detect these intrusion attempts so that action may be taken to repair the damage later. This field of research is called Intrusion Detection.

A simple firewall can no longer provide enough security as in the past. Today's corporations are drafting intricate security policies whose enforcement requires the use of multiple systems, both proactive and reactive (and often multi-layered and highly redundant). The premise behind intrusion detection systems is simple: Deploy a set of agents to inspect network traffic and look for the "signatures" of known network attacks. However, the evolution of network computing and the awesome availability of the Internet have complicated this concept somewhat. With the advent of Distributed Denial of Service (DDOS) attacks, which are often launched from hundreds of separate sources, the traffic source no longer provides reliable temporal clues that an attack is in progress. Worse yet, the task of responding to such attacks is further complicated by the diversity of the source systems, and especially by the geographically distributed nature of most attacks.

Intrusion detection techniques while often regarded as grossly experimental, the field of intrusion detection has matured a great deal to the point where it has secured a space in the network defense landscape alongside firewalls and virus protection systems. While the actual implementations tend to be fairly complex, and often proprietary, the concept behind intrusion detection is a surprisingly simple one: Inspect all network activity (both inbound and outbound) and identify suspicious patterns that could be evidence of a network or system attack.

III CLASSIFICATION OF IDS

One of the main approaches of IDS, namely anomaly detection is based on the assumption that an attack on a computer system will be noticeably different from normal system activity, and an intruder will exhibit a pattern of behavior different from that of the normal user. In the second leading approach, misuse detection, a collection of known intrusion techniques is kept in a knowledge base, and intrusions are detected by searching through the knowledge base.

A. Anomaly Detection

Anomaly detection techniques assume that all intrusive activities are necessarily anomalous. This means that if we could establish a "normal activity profile" for a system, we could, in theory, flag all system states varying from the established profile by statistically significant amounts as intrusion attempts. However, if we consider that the set of intrusive activities only intersects the set of anomalous activities instead of being exactly the same, we find a couple of interesting possibilities: (1) Anomalous activities that are not intrusive are flagged as intrusive. (2) Intrusive activities that are not anomalous result in false negatives (events are not flagged intrusive, though they actually are). This is a dangerous problem, and is far more serious than the problem of false positives [3]. The main issues in anomaly detection systems thus become the selection of threshold levels so that neither of the above 2 problems is unreasonably magnified, and the selection of features to monitor. Anomaly detection systems are also computationally expensive because of the overhead of keeping track of, and possibly updating several system profile metrics. A block diagram of a typical anomaly detection system is shown in Figure below.

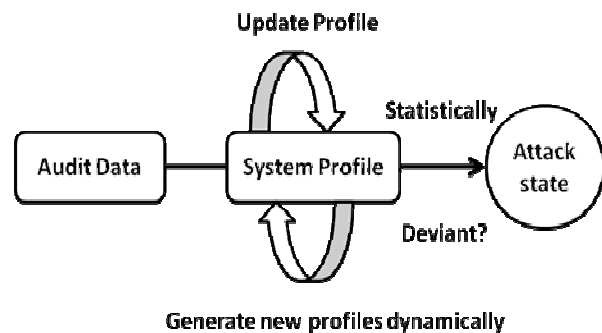


Figure 2. A Typical Anomaly-based Detection System.

B Misuse Detection

The concept behind misuse detection schemes is that there are ways to represent attacks in the form of a pattern or a signature so that even variations of the same attack can be detected [4]. This means that these systems are not unlike virus detection systems -- they can detect many or all *known* attack patterns, but they are of little use for as yet unknown attack methods. An interesting point to note is that anomaly detection systems try to detect the complement of "bad" behavior. Misuse detection systems try to recognize known "bad" behavior. The main issues in misuse detection systems are how to write a signature that encompasses *all* possible

variations of the pertinent attack, and how to write signatures that do not also match non-intrusive activity. A block diagram of a typical misuse detection system is shown in Figure below.

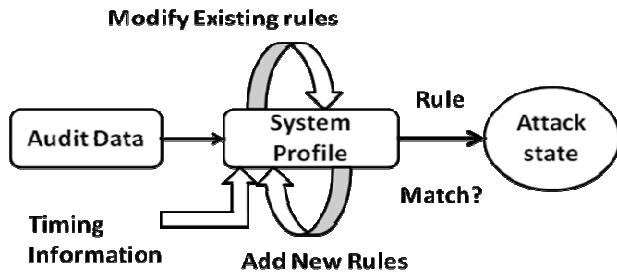


Figure 3. A Typical Misuse-based Detection System

IV SNORT ARCHITECTURE

Snort is primarily a rule-based IDS, however input plug-ins are present to detect anomalies in protocol headers. Snort is a very powerful tool and is known to be one of the best IDS on the market even when compared to commercial IDS. Snort uses rules stored in text files that can be modified by a text editor. Rules are grouped in categories. Rules belonging to each category are stored in separate files. These files are then included in a main configuration file called `snort.conf`. Snort reads these rules at the start-up time and builds internal data structures or chains to apply these rules to captured data.

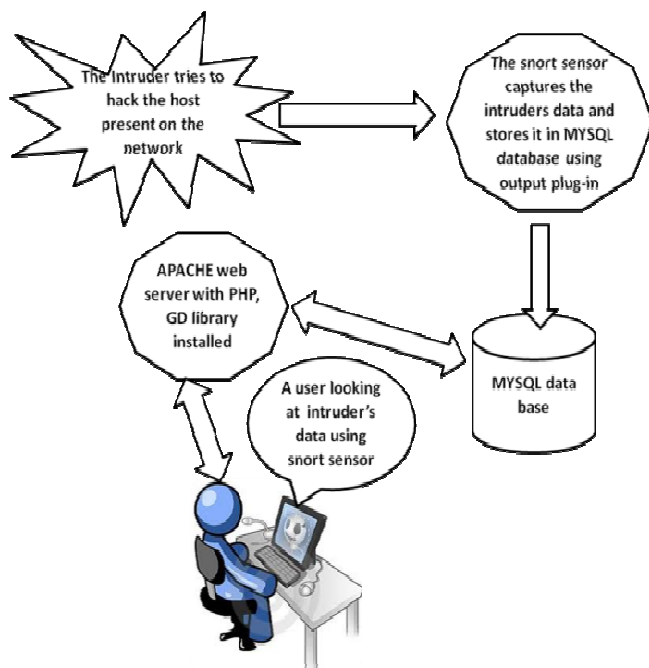


Figure 4. Block diagram of a complete network intrusion detection system.

A. How Is Snort Different From tcpdump?

The major feature that Snort has which tcpdump does not is packet payload inspection. Snort decodes the application layer of a packet and can be given rules to collect traffic that has specific data contained within its application layer. One powerful feature that Snort and tcpdump share is the capability to filter traffic with Berkeley Packet Filter (BPF) commands. This allows traffic to be collected based upon a variety of specific packet fields. For example, both tools may be instructed via BPF commands to process TCP traffic only

[5]. While tcpdump would collect all TCP traffic, Snort can utilize its flexible rules set to perform additional functions, such as searching out and recording only those packets that have their TCP flags set a particular way or containing web requests that amount to CGI vulnerability probes.

B. Components

Snort can be divided into five major components that are each critical to intrusion detection.

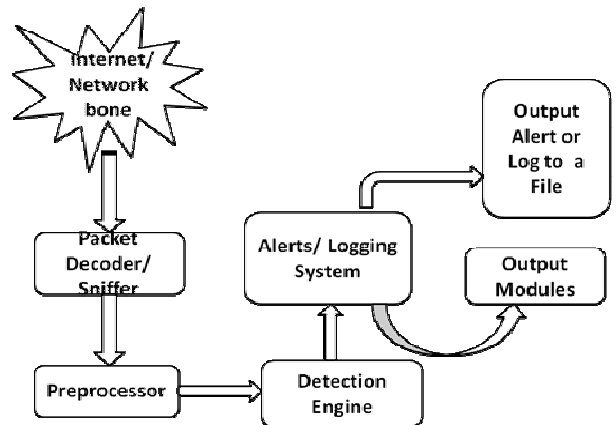


Figure 5. Various components used in snort.

The first is the packet capturing mechanism. Snort relies on an external packet capturing library (libpcap) to sniff packets. After packets have been captured in a raw form, they are passed into the packet decoder. The decoder is the first step into Snort's own architecture. The packet decoder translates specific protocol elements into an internal data structure. After the initial preparatory packet capture and decode is completed, traffic is handled by the preprocessors. Any number of pluggable preprocessors either examines or manipulates packets before handing them to the next component: the detection engine. The detection engine performs simple tests on a single aspect of each packet to detect intrusions. The last component is the output plugins, which generate alerts to present suspicious activity to you.

To get packets into the preprocessors and then the main detection engine, some prior labor must first occur. Snort has no native packet capture facility yet; it requires an external packet sniffing library: libpcap. Libpcap was chosen for packet capture for its platform independence. It can be run on every popular combination of hardware and OS; there is even a Win32 port—winpcap. Because Snort utilizes the libpcap library to grab packets off the wire, it can leverage libpcap's platform portability and be installed almost anywhere. Using libpcap makes Snort a truly platform-independent application.

Using libpcap is not the most efficient way to acquire raw packets. It can process only one packet at a time, making it a bottleneck for high-bandwidth (1Gbps) monitoring situations. In the future Snort will likely implement packet capture libraries specific to an OS, or even hardware. There are several methods other than using libpcap for grabbing packets from a network interface card. Berkeley Packet Filter (BPF), Data Link Provider Interface (DLPI), and the SOCK_PACKET mechanism in the Linux kernel are other tools for grabbing raw packets.

A *raw packet* is a packet that is left in its original, unmodified form as it had traveled across the network from client to server. A raw packet has all its protocol header information left intact and unaltered by the operating system. Network applications typically do not process raw packets; they depend on the OS to read protocol information and properly forward payload data to them. Snort is unusual in this sense in that it requires the opposite: it needs to have the packets in their raw state to function. Snort uses protocol header information that would have been stripped off by the operating system to detect some forms of attacks.

An illustration of Snort's packet processing is given in Fig. 6 [6].

Snort is a rule-based network intrusion detection system (NIDS). It has a flexible rule defining language that lets anyone to change existing rules or adding new rules to the IDS. Every rule consists of two logical parts: the rule header and rule options. Rule header has five sections; rule actions (action to be taken when an intrusion is detected), end-to-end source and destination information (source and destination IP addresses and port numbers depending on the protocol), and direction of traffic and protocol type (TCP, UDP, or ICMP).

Rule options consist of various conditions that help deciding whether the mentioned misuse operation has occurred or not. A sample Snort rule is shown in Fig. 7. The first field of every rule is the action field. This field can have the following values: log, alert, pass, activate, or dynamic. When the input value matches the criteria, these actions are taken as a response. The selected action in Fig. 7 is "alert". This states that, if an entry matches with the mentioned criteria, an alert will be created. The next field holds the protocol information. Valid values of this field can be TCP, UDP, or ICMP. The protocol in our example is TCP. The third and the fourth fields hold source addresses; the first part stands for IP address and the second part is the port number. If this field has the value "any any", it means that the packets may be originating from any IP address and any TCP port. In the case where protocol value is ICMP, no port value is used as this field is meaningful for only TCP and UDP.

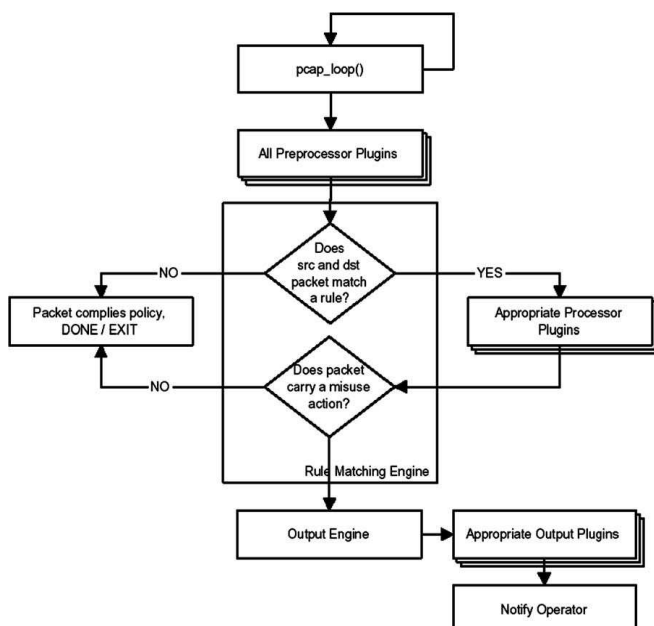


Figure 6. Snort packet flow

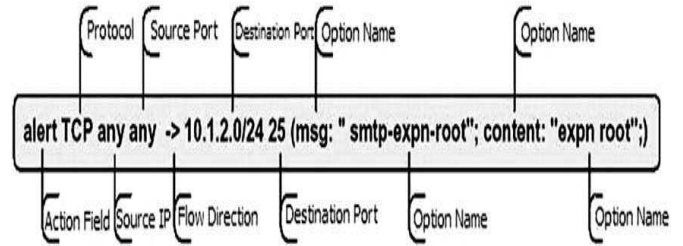


Figure 7. Rule Structure

The fifth field shows the flow direction of the information. The sixth and the seventh fields hold destination addresses. The example destination IP address is given as 10.1.2.0/24, which matches all the IP addresses in a C class network. In this example, TCP destination port is set as 25. Port 25 is used for simple mail transfer protocol (SMTP).

Following the destination address, there is an options list written in parenthesis. Every option consists of an option name, an option value if exists and a semicolon indicating the end of that option. The first option shown in Fig. 7 is "msg" and it is used to state the action message. "Content" is the second option and states a template-matching criterion. In the sample shown in Fig. 7, "expn root" string is searched for in the input data. If a match is found in TCP data segment, the condition is met.

All of the below criteria must be met in order to make the sample rule shown in Fig. 7 produce an alert:

- The entry must be a TCP packet.
- The entry may be originating from any IP address and any TCP port.
- The entry must destine the 10.1.2.0/24 network and port number 25 of the computer located in this network.
- The entry must include "expn root" string.

V PERFORMANCE OF SNORT

The intrusion detection system (IDS) offers intelligent protection of networked computers or distributed resources much better than using fixed-rule firewalls. Snort was designed to fulfill the requirements of a prototypical lightweight network intrusion detection system. It has become a flexible and highly capable system that is in use around the world. Snort is a signature -based IDS which detects attacks by matching against a database of known attacks. The signatures are manually constructed by security experts analyzing previous attacks. The collected signatures are used to match with incoming traffic to detect intrusions. The snort maintains a database of attack signatures and uses an alert signal .If the rule matches, the signal fires. Snort is a fully capable alternative to commercial intrusion detection systems in places where it is cost inefficient to install full featured commercial systems. If you implement a special hardware called hybrid architecture (ADS with snort), thereby increase the preprocessing ability and to achieve higher detection accuracy.

VI A NEW HYBRID IDS

Our research has been to design a pre-processor to allow detection of anomalies that Snort turn into a hybrid system. This system, named Hybrid IDS meets the following requirements:

1. It models the network traffic at high level.
2. It stores the information in a database in order to model the normal behavior of the system.
3. It is totally configurable and allows adjusting the sensitivity of the system to prevent false alarms.
4. It has two operation phases: training and anomaly detection.

5. It is complemented with a website that allows the user can administrate and observe the network performance.

Snort has been extended by adding an anomaly detection pre-processor which access to a database MySQL where it is centralized the system configuration, statistical data and anomalies detected by the system. The system is complemented by a website that displays the system status (network traffic, detected anomalies, etc) and that also allows to configure the system easily.

A. The Anomaly Detection Pre-processor

The anomaly detection module is responsible of recording all the abnormal activity. Figure shows the general scheme of the anomaly detection module using two different operation modes: training mode and anomaly detection mode. Using the training mode the system records in a database the network traffic considered as normal and expected. Later, a profile of this network activity is automatically created, and the anomaly detection module stores in the database the

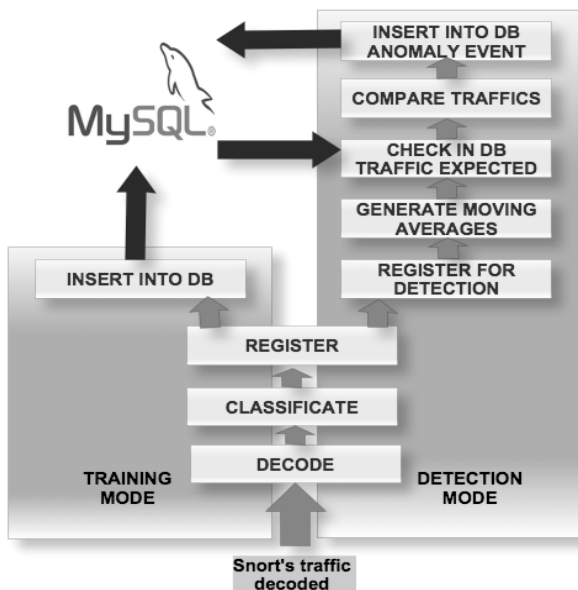


Figure 8. Anomaly Detection Processor

Both operation modes share the same functionality. When the pre-processor of Snort receive a package, it is classified according to its class (if the package is primary/secondary and if the package belongs to a network server or a client), and it stores the vector-class package, i.e.: the system is recording and counting the network traffic. When the system is in training mode, it stores the recorded information in the database, and later it obtains a profile of the normal activity. The information stored in the database is used when the system is in detection mode. Daily and each time the system is executed, the activity profiles of the most active clients and servers in the network are loaded from the database. Therefore, as the expected traffic is recorded in the database, it is compared with the real traffic passing through the

network. If it is detected a deviation in the traffic higher than a certain percentage, it means that something abnormal is happening, and an incidence of abnormality is registered by the system. It is remarked that the system must compare the received traffic with the activity previously stored in training mode. With this aim, several techniques have been implemented such as statistical methods [7], expert systems [8], data mining [9], etc.

VII HYBRID INTELLIGENT IDS ARCHITECTURE

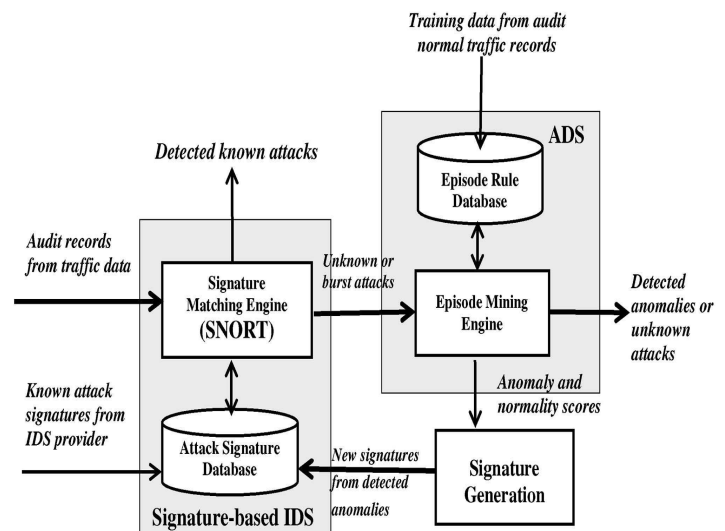


Figure 9. Hybrid Intelligent IDS architecture

Anomaly-based systems are supposed to detect unknown attacks. These systems are often designed for offline analysis due to their expensive processing and memory overheads. Signature-based system leverages manually characterized attack signatures to detect known attacks in real-time traffic. The HIDS illustrated in Fig. 9 integrates the flexibility of ADS with the accuracy of a signature-based SNORT. The SNORT is connected in cascade with the custom-designed ADS. These two subsystems join hands to cover all traffic events initiated by both legitimate and malicious users.

By 2004, SNORT has accumulated more than 2,400 attack signatures in its database [10]. In HIDS operations, the first step is to filter out the known attack traffic by SNORT through signature matching with the database. The remaining traffic containing unknown or burst attacks is fed to the episode-mining engine to generate frequent episode rules with different levels of support threshold. This leveling allows the detection of some rare episodes, declared as anomalies. The frequent episodes are compared with precompiled frequent episodes from normal traffic. The episodes that do not match the normal profiles or match them with unusually high frequency are labeled as anomalous. The anomalous episodes are used to generate signatures which capture the anomalous behavior using a weighted frequent item set mining scheme. These signatures are then added to the SNORT database for future detection of similar attacks. Unknown, burst, or multi connection attacks are detectable by ADS. The signature generation unit bridges two detection subsystems in the shaded boxes. This unit characterizes the detected anomalies and extracts their signatures. We built ADS by using the FER mining mechanisms. The new HIDS detects many novel attacks hidden in common Internet services, such as telnet, http, ftp, smtp, e-mail, authentication, and so forth. The HIDS deployment appeals

particularly to protect network-based clusters of computers, resources inside internal networks (intranets), and computational grids.

VIII DATA MINING SCHEME FOR NETWORK ANOMALY DETECTION

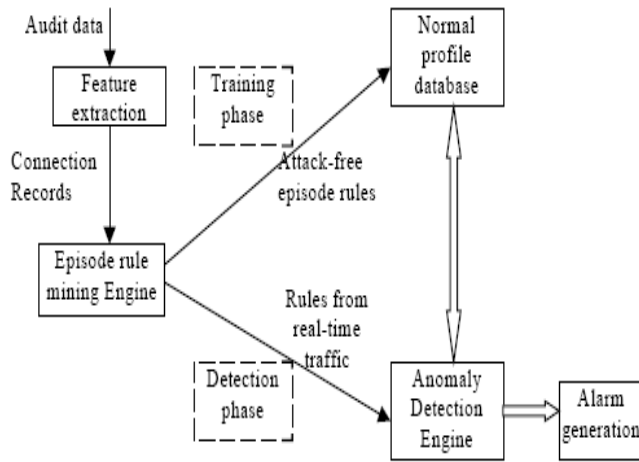


Figure 10. Datamining architecture for anomaly-based intrusion detection

A. Internet Connection Episode Rules

New datamining techniques are developed for generating frequent episode rules of traffic events. These episode rules are used to distinguish anomalous sequences of TCP, UDP, or ICMP connections from normal traffic.

In order to distinguish between intrusive and normal network traffic, new datamining algorithms are developed to generate frequent episode rules (FER) [11] from audit Internet records. An episode is represented by a sequence of Internet connections.

The tasks of datamining are described by association rules or by frequent episode rules. An association rule is aimed at finding interesting intra-relationship inside a single connection record. The FER describes the inter-relationship among multiple connection records in a sequence. The FER is more powerful to characterize traffic episodes than using the association rules alone.

Frequent Episode Rules: In general, an FER is expressed by the expression:

$$L_1, L_2, \dots, L_n \rightarrow R_1, \dots, R_m (c, s, \text{window}) \text{ -----(1)}$$

where L_i ($1 \leq i \leq n$) and R_j ($1 \leq j \leq m$) are ordered item sets in a traffic record set T . We call L_1, L_2, \dots, L_n the LHS (left hand side) episode and R_1, \dots, R_m the RHS (right hand side) episode of the rule. Note that all item sets are sequentially ordered, that is $L_1, L_2, \dots, L_n, R_1, \dots, R_m$ must occur in the ordering as listed. However, other item sets could be embedded within our episode sequence. We define the support and confidence of rule (1) by the following two expressions:

$$S = \text{support} (L_1 \cup L_2 \dots \cup R_1 \dots \cup R_m) \text{ greater than or equal to } s_0 \text{ -----(2)}$$

$$C = \frac{\text{support} (L_1 \cup L_2 \dots \cup R_1 \dots \cup R_m)}{\text{Support} (L_1 \cup L_2 \cup \dots \cup L_n)} \text{ Greater than equal to } c_0 \text{ -----(3)}$$

An example FER is given below for a sequence of network events:

(Service = authentication) \rightarrow (service = SMTP) (service = SMTP) (0.6, 0.1, 2 sec)

This rule specifies an authentication event. If the authentication service is requested at time t , there is a confidence level of $c = 60\%$ that two SMTP services will follow before the time $t + w$, where the event window $w = 2$ sec. The support of 3 traffic events (service = authentication), (service = SMTP), (service = SMTP) accounts for 10% of all network connections.

Here we consider the minimal occurrence introduced by Mannila et al [12] of the episode sequence in the entire traffic stream. The support value s is defined by the percentage of occurrences of the episode within the parentheses out of the total number of traffic records audited. The confidence level c is the probability of the minimal occurrence of the joint episodes out of the LHS episode. Both parameters are lower bounded by s_0 and c_0 , the minimum support value and the minimum confidence level, respectively. The window size is an upper bound on the time duration of the entire episode sequence.

The traffic connections on both sides of a FER need not be disjoint in an episode sequence of events. Episode rules can be used to characterize attacks. The SYN flood attack is specified by the following episode rule:

Where the event (service = http, flag = S0) is an association. Flag **S0** signals only the SYN packet being seen in a particular connection. The combination of associations and FERs reveals useful information on normal and intrusive behaviors. These rules can be applied to build IDS to defend against both known and unknown attacks.

IX INTRUSION DETECTION VIA FUZZY LOGIC AND DATA MINING

Data mining is one of the hot topics in the field of knowledge extraction from database. Data mining is used to automatically learn patterns from large quantities of data.

Mining can efficiently discover useful and interesting rules from large collection of data. Fuzzy logic provides a powerful way to categorize a concept in an abstract way. The advantage of fuzzy logic is that it allows representation of overlapping categories. We are combining techniques from fuzzy logic and data mining for anomaly detection and it helps us to create more abstract patterns.

A. Fuzzy Logic

Fuzzy concepts derive from fuzzy phenomena that commonly occur in the natural world. For instance "rain" is a fuzzy statement of "Today raining heavily" since there is no clear boundary between "rain" and "heavy rain". In intrusion detection suppose we want to write a rule as given below we need a reason about a quantity such as the number of different destination IP addresses in the last 2 seconds.

IF the number of different destination addresses during the last n seconds was high

THEN an unusual situation exists.

Fuzzy logic, which is utilized, is a superset of conventional logic that has been extended to handle the concept of partial truth, which lies between completely true and completely false.

B. Data Mining Methods

Data mining methods are used to automatically discover new patterns from a large amount of data. Association rules have

been successfully used to mine audit data to find normal patterns for anomaly intrusion detection [13].

C. Association Rules

The association rule mechanism proposed by Agrawal is a most popular tool. Agrawal and Srikant [13] developed the Apriori Algorithm for mining association rules. The Apriori Algorithm needs confidence (to represent minimum confidence) and support (to represent minimum support). These two values determine the degree of association that must hold.

D. Fuzzy Association Rules

To efficiently use the Apriori algorithm of Agrawal and Srikant [14] for mining association rules, quantitative variables should be partitioned into discrete categories. This leads to "sharp boundary problem" in which a very small change in value causes an abrupt change in category. To address this problem fuzzy association rules was developed by Kuok, Fu and Wong [15]. We modify the algorithm [15] by introducing a normalization factor to ensure that every transaction is counted only one time

X FUTURE WORK

We suggest the following issue for continued research and development effort i.e. distributed environment.

A Distributed HIDS consists of several IDS over a large network (s), all of which communicate with each other, or with a central server that facilitates advanced network monitoring. Distributed HIDS also helps to control the spreading of worms, improves network monitoring and incident analysis, attack tracing and so on. It also helps to detect new threats from unauthorized users, back-door attackers and hackers to the network across multiple locations, which are geographically separated. These systems require the audit data collected from different places to be sent to a central location for an analysis. Since the amount of audit data that an IDS needs to examine is very large even for a small network. So reducing the data is important in this process. Finally it calculates the performance and compared with previously implemented techniques.

XI CONCLUSION

Signature-based systems can only detect attacks that are known before whereas anomaly-based systems are able to detect unknown attacks. Anomaly-based IDSs make it possible to detect attacks whose signatures are not included in rule files.

Both SNORT and ADS subsystems have low processing overhead. The integration of ADS with snort has upgraded the SNORT detection rate by 40 percent with less than a 1 percent increase in false alarms. Generating more signatures by ADS will further enhance the overall performance of the hybrid IDS. In order to improve the detection rate, we implement a framework for Distributed Intrusion Detection Systems (DIDS) with a focus on improving the intrusion detection performance by reducing the input features. Finally with the increasing incidents of cyber attacks, building an effective intrusion detection models with good accuracy and real-time performance are essential. This field is developing continuously. More data mining techniques should be investigated and their efficiency should be evaluated as intrusion detection models

REFERENCES

- [1] Heady, R., Luger, G., Maccabe, A. & Servilla, M. 1990. The architecture of a network level network intrusion detection system. Technical report CS90-20, Department of Computer Science, University of New Mexico.
- [2] Tao Peng, Christopher Leckie, Kotagiri Ramamohanarao. 2007, Information sharing for distributed intrusion detection systems Journal of Network and Computer Applications 30 (2007) 877-899.
- [3] Sundaram, A. 1996. An introduction to intrusion detection. Crossroads: The ACM student magazine, 2(4), April.
- [4] Rachid Beghdad, 2004, Modelling and solving the intrusion detection problem in computer networks, Computers & Security (2004) 23, 687-696.
- [5] tcpdump, Van Jacobson, Craig Leres and Steven McCanne, Lawrence Berkeley National Laboratory, 1991, <http://www.nrg.ee.lbl.gov/>.
- [6] Dayiog'lu B. Use of passive network mapping to enhance network intrusion detection. Thesis (Master), The Graduate School of Natural and Applied Sciences of the Middle East Technical University; 2001
- [7] Ye, N., Emran, S.M., Li, X., Chen, Q.: Statistical process control for computer intrusion detection. In: DARPA Information Survivability Conference & Exposition II, DISCEX 2001 (2001)
- [8] Denning, D.E.: An Intrusion-Detection Model. IEEE Transactions on Software Engineering 13(2), 222-232 (1987)
- [9] Barbara, D., Wu, N., Jajodia, S.: Detecting novel network intrusions using Bayes estimators. In: Proceedings of First SIAM Conference on Data Mining, Chicago, IL (2001)
- [10] B. Casewell and J. Beale, SNORT 2.1, Intrusion Detection, second ed. Syngress, May 2004.
- [11] H. Mannila, H. Toivonen, and A. I. Verkamo. "Discovery of Frequent Episodes in Event Sequences", *Data Mining and Knowledge Discovery*, 1(3), 1997. [12] H. Mannila and H. Toivonen. "Discovering Generalized Episodes using Minimal Occurrences", *Proc. of the Second Int'l Conf. on knowledge discovery and datamining*, Portland, Oregon, August, 1996.
- [13] Lee, W., S. Stolfo and K. Mok 1998 "Mining audit data to build intrusion detection models". *Fourth international conference on knowledge discovery and data mining*, New York August 1998.
- [14] Agrawal, R., and R. Srikant 1994 "Fast algorithms for mining association rules 20" *h international conference on very large databases* September 1994.
- [15] Kuok, C., A. Fu and M. Wong "Mining fuzzy association rules in databases" *SIGMOD Record* 17 (1) 41-46.

AUTHOR S PROFILE

Sravan Kumar Jonnalagadda Pursued his M.Tech in Computer Science at Acharya Nagarjuna University. He received his Bachelor degree in Computer Science in 2008. Now, he is an Assistant Professor in department of Information Technology in DMS SVH College of engineering, Machilipatnam (India). His current research interest includes Network Security and Software Engineering.

Subha Sree Mallela Pursued her M.Tech in Computer Science and Engineering at Acharya Nagarjuna University. She received her Bachelor degree in Computer Science in 2008. Now, she is an Assistant Professor in department of Computer Science in DMS SVH College of engineering, Machilipatnam (India). Her current research interest includes Network Security, Software Engineering and Aspect Oriented Programming.

A Matlab Implementation Of The Back-Propagation Approach for Reusability of Software Components

Meenakshi Sharma¹
Sri Sai College Of Engg. & Tech., Pathankot²
HOD CSE²
Mss.s.c.e.t@gmail.com

Priyanka Kakkar²
Sri Sai College Of Engg. & Tech., Pathankot² Mtech CSE
4th sem²
pinkudiya@yahoo.co.in

Dr. Parvinder Sandhu³
Rayat and Bhara College OF Engg. And Tech.³
HOD CSE³

Abstract- Before the age of the computer, there were many mathematical problems that humans could not easily solve, or more precisely (and this distinction is extremely important) humans were too slow in solving. Computers enabled these often simple but slow and tedious tasks to be performed quickly and accurately. The first problems solved with computers were calculating equations to resolve important physical problems, and later displaying a nice GUI, making word processors and so on. However, there are many common tasks which are trivial for humans to perform (without even any conscious effort) yet which are extremely difficult to formulate in a way that a computer may easily solve.

Keywords:- Reusability, JISC, GUI, back-propagation algorithm.

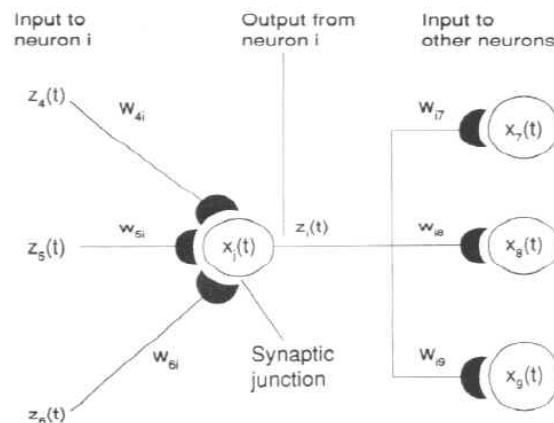
I. INTRODUCTION (HEADING 1)

Generalising the Widrow-Hoff learning rule to multiple-layer networks and non-linear[1][2] differentiable transfer functions created back propagation. Input vectors and the corresponding output vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined by you.

Networks with biases, a sigmoid layer, and a linear output layer are capable of approximating any function[3] with a finite number of discontinuities. Standard back propagation is a gradient descent algorithm, as is the Widrow-Hoff learning rule. There are a number of variations [4][5][6] on the basic

algorithm, which are based on other standard optimisation techniques, such as conjugate gradient and Newton methods.

The term back propagation refers to the process by which derivatives of network error, with respect to network weights and biases, can be computed can compute derivatives of network error. This process can be used with a number[7] of different optimisation strategies. The architecture of a multilayer network is not completely constrained by the problem[2] to be solved. The number of inputs to the network is constrained by the problem, and the number of neurons in the output layer is constrained by the number of outputs required by the problem. However, the number of layers between network inputs and the output layer and the sizes of the layers are up to the user.



A unit in the output layer determines its activity by following a two step procedure.

- First, it computes the total weighted input x_j , using the formula:

$$X_j = \sum_i y_i W_{ij}$$

where y_i is the activity level of the i th unit in the previous layer and W_{ij} is the weight of the connection between the i th and the j th unit.

- Next, the unit calculates the activity y_j using some function of the total weighted input. Typically we use the sigmoid function:

$$y_j = \frac{1}{1 + e^{-x_j}}$$

Once the activities of all output units have been determined, the network computes the error E , which is defined by the expression:

$$E = \frac{1}{2} \sum_i (y_i - d_i)^2$$

where y_j is the activity level of the j th unit in the top layer and d_j is the [8] desired output of the j th unit.

The back-propagation algorithm consists of four steps:

1. Compute how fast the error changes as the activity of an output unit is changed. This error derivative (EA) is the difference between the actual and the desired activity.

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

2. Compute how fast the error changes as the total input received by an output unit is changed. This quantity (EI) is the answer from step 1 multiplied by the rate at which the output of a unit changes as its total input is changed.

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{dy_j}{dx_j} = EA_j y_j (1 - y_j)$$

3. Compute how fast the error changes as a weight on the connection into an output unit is changed. This quantity (EW) is the answer from step 2 multiplied by the activity level of the unit from which the connection emanates.

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial W_{ij}} = EI_j y_i$$

4. Compute how fast the error changes as the activity of a unit in the previous layer is changed. This crucial step[9] allows back propagation to be applied to multilayer networks. When the activity of a unit in the previous layer changes, it affects the activities of all the output units to which it is connected. So to compute the overall effect on the error, we add together all these separate effects on output units. But each effect is simple to calculate. It is the answer in step 2 multiplied by the weight on the connection to that output unit.

$$EA_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial y_i} = \sum_j EI_j W_{ij}$$

By using steps 2 and 4, we can convert the EAs of one layer of units into EAs for the previous layer. This procedure can be repeated to get the EAs for as many previous layers as desired. Once we know the EA of a unit, we can use steps 2 and 3 to compute the [10]EWs on its incoming connections.

2. Methodology:

A. Pilot interviews

Early interviews were conducted with three projects on a pilot basis to allow us to refine our interview structure and to develop and modify our explicit methodology.

B. Detailed Evaluation Criteria document

In close consultation with JISC programme staff, we drafted, amended, and agreed this document, which defines the software evaluation criteria to be used to evaluate the 22 DeLeTools projects. This document, which is available on the JISC website, is closely related to the accompanying Evaluation Methods and Tools document mentioned below: the criteria document describes the tasks to be done and what criteria we use to [11] approach them, and the methods document describes how we should do it.

C. Detailed Evaluation Methods and Tools document

Similarly, we prepared this document, which details the methods used to assess, test, and evaluate the software outputs of the 22 projects in the DeLeTools strand of the JISC e-Learning Programme. Using the Software Quality Evaluation

Criteria[12] Document as a high level guide, the Detailed Evaluation Methods and Tools document outlines the specific methods proposed and references the criteria that these methods address. The latter document states:

Our approach is based not only on the quality standards to which JISC require the 22 projects to adhere and the individual standards which have been identified by the projects themselves, but also the maintainability, extendibility and the general robustness of the 22 projects' code and software. Code review will be a key focus of the evaluation of each project. Other evaluation and testing methods will vary according to the project and its outputs.

D. Summary of methods used

We used the following methods:

1. Brief Online Study.
2. Face to face interview with project staff and follow up by email and phone (full accounts of interviews appear in Appendix).
3. Selective checking of sections of code (per-project summary of code checking results appears below). [13]
4. Lab testing the software on various platforms and in various conditions (per-project summary of test results appears below – full results have been made available to projects and to programme staff).
5. Usability walkthrough and application of heuristics (per-project summary of usability results appears below – full results have been made available to projects and to programme staff).
6. Brief and selective desk audit of project documentation, version tracking, and testing procedures (the results of the audit were fed into the software evaluation and gave the testers insight into which areas of the tool might need particular[14] attention, also alerting the testing team to any issues and bugs prior to testing and informing some project-specific questions asked during the interviews).

II. CRITERIA USED IN EVALUATION

The criteria used in the evaluation[15] and a brief summary of the measures or "questions asked" appears below.

A. Match of actual software output to planned output Measures

Are the outputs materially different from those planned? Are they so delayed that the project will not, or is unlikely to, deliver those outputs?

B. Use of quality plan Measures

Has the quality plan been completed and followed? Has it been used to aid implementing the JISC Software Quality Assurance Policy? Has it been updated? Has it been used in creative[16] and unforeseen ways?

C. Compliance with the JISC Open Source Policy Measures

The (draft) JISC Open Source Policy (May 2004) itself details the necessary areas of compliance.

D. Compliance with Open Standards Measures

Have the Open Standards outlined in the quality plan been used? If not, have other Open Standards been used?

E. Quality control procedures Measures

Are there any documented formal or informal quality control procedures? Is there evidence of them being used? Can we reproduce testing procedures? Are issues and changes being documented?

F. Project specific documentation Measures

In addition to the project plan and quality plan, we asked each project to give us access to:

- The technical specification of their product
- a summary of their testing procedures
- a copy of their test plan
- reports from version tracking software
- code documentation. [17]

It was anticipated that there would be little or no end-user documentation, and that few projects would have a complete technical specification document.

3.Results:

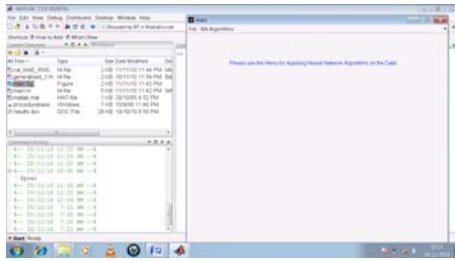


Figure 6.1- Main Menu

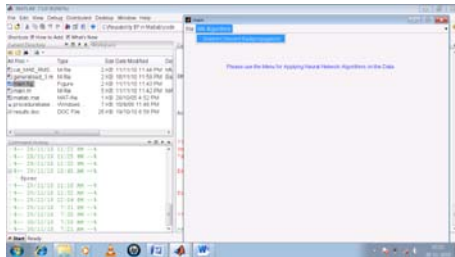


Figure6.2- choosing Option

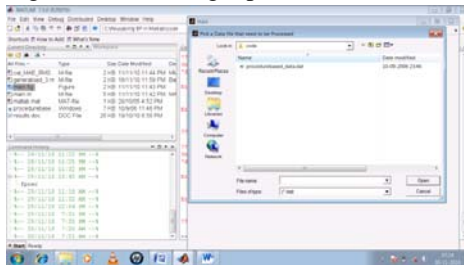


Figure6.3- Selecting Source File

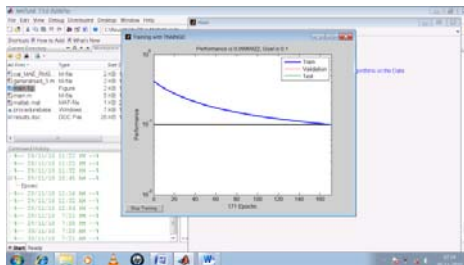


Figure6.4- Final result.

4.CONCLUSION

Despite some reservations stated above about the ESI questionnaire, the survey has achieved its goal of providing benchmark information about adoption of best practice by software developers in Queensland. At the time of writing, about 40 responses have been received by SEA in Western Australia and the WA findings will provide interesting comparisons against those from Queensland. The proportion and best practice adoption level of COTS developers will be of particular interest in relation to the Queensland findings.

Although all the follow-up meetings for the Process Improvement Program[18] have not yet been completed, feedback from the participant organisations has been positive.

Thompson (1994) is concerned that technology transfer of appropriate software engineering practices is inhibited by management and developers because they often have poor attitudes towards change. He advises that the software development community needs a[19] much better understanding of the practices, their use and potential benefits. It is hoped that through the publication of these success stories from the Process Improvement Program, local developers will appraise such information in an impartial manner and adopt best practice willingly and with enthusiasm.

A valuable outcome of the Process Improvement Program is the development of the RAPID evaluation tool which provides a realistic option to very small development organisations which traditionally lack the resources to undertake full-blown software process assessments. With the Australian software community dominated by very small organisations (88 percent have less than five staff), this program may provide valuable opportunities for such developers to evaluate and improve their processes, thereby achieving success in domestic and global markets.

Funding has been made available to provide full SPICE assessments and mentoring for two of the Process Improvement Program participants. It is anticipated that case studies, in a similar format to the SPIRE (Software Process Improvement in Regional Europe) Case Studies will be compiled and published. The Federal Government's support for software incubators, such as the one operating at SEA Qld, promises an opportunity for small start-up companies to overcome traditional[20] resource limitations. These steps may facilitate the achievement of the goal of improved global competitiveness for the Australian software industry.

5.REFERENCES

- [1] Sandhu P. S. and Singh H., "A Reusability Evaluation Model for OO-Based Software Components", International Journal of Computer Science, vol. 1, no. 4, 2006, pp. 259-264.
- [2] Neural Networks at Pacific Northwest National Laboratory <http://www.emsl.pnl.gov:2080/docs/cie/neural/neural.homepage.html>
- [3] Artificial Neural Networks in Medicine http://www.emsl.pnl.gov:2080/docs/cie/techbrief/NN.techbrief_.ht
- [4] A Novel Approach to Modelling and Diagnosing the Cardiovascular System http://www.emsl.pnl.gov:2080/docs/cie/neural/papers2/keller_wcnn95.abs.html
- [5] Electronic Noses for Telemedicine <http://www.emsl.pnl.gov:2080/docs/cie/neural/papers2/keller.c>

cc95.abs.html

- [6] G. Caldiera and V. R. Basili, Identifying and Qualifying Reusable Software Components, *IEEE Computer*, (1991) 61-70.
- [7] W. Tracz, A Conceptual Model for Mega programming, *SIGSOFT Software Engineering Notes*, 16(3, July 1991) 36-45.
- [8] Stephen R. Schach and X. Yang, Metrics for targeting candidates for reuse: an experimental approach, *ACM, (SAC 1995)* 379-383.
- [9] J. S. Poulin, *Measuring Software Reuse—Principles, Practices and Economic Models* (Addison-Wesley, 1997).
- [10] W. Humphrey, *Managing the Software Process*, SEI Series in Software Engineering (Addison-Wesley, 1989).
- [11] L. Sommerville, *Software Engineering*, 4th edn. (Addison-Wesley, 1992).
- [12] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, 5th edn. (McGraw-Hill, 2005).
- [13] G. Boetticher and D. Eichmann, A Neural Network Paradigm for Characterizing Reusable Software, *Proc. of the 1st Australian Conference on Software Metrics* (18-19 November 1993).
- [14] S. V. Kartalopoulos, *Understanding Neural Networks and Fuzzy Logic-Basic Concepts and Applications* (IEEE Press, 1996)153-160.
- [15] Parvinder Singh Sandhu and Hardeep Singh, "Software Reusability Model for Procedure Based Domain-Specific Software Components", *International Journal of Software Engineering & Knowledge Engineering (IJSEKE)*, Vol. 18, No. 7, 2008, pp. 1–19.
- [16] Parvinder Singh Sandhu and Hardeep Singh, "Automatic Quality Appraisal of Domain-Specific Reusable Software Components", *Journal of Electronics & Computer Science*, vol. 8, no. 1, June 2006, pp. 1-8.
- [17] Parvinder Singh Sandhu and Hardeep Singh, "A Reusability Evaluation Model for OO-Based Software Components", *International Journal of Computer Science*, vol. 1, no. 4, 2006, pp. 259-264.
- [18] Parvinder Singh Sandhu and Hardeep Singh , "Automatic Reusability Appraisal of Software Components using Neuro-Fuzzy Approach, *International Journal Of Information Technology*, vol. 3, no. 3, 2006,pp. 209-214.
- [19] Parvinder S. Sandhu and Hardeep Singh, "A Fuzzy Based Approach for the Prediction of Quality of Reusable Software Components", *IEEE 14th International Conference on Advanced Computing & Communications (ADCOM 2006)*, NIT Suratkal, Dec. 20 – 23, 2006, pp. 761-764.
- [20] Parvinder S. Sandhu and Hardeep Singh, "A Neuro-Fuzzy Based Software Reusability Evaluation System with Optimized Rule Selection", *IEEE 2nd International*

Conference on Emerging Technologies (IEEE ICET 2006), Peshawar, Pakistan, Nov. 13-14, 2006, pp. 664-669.

Decision tree Induction Algorithm for Classification of Image Data

Kesari verma

Department of Computer Science
National Institute of Technology
Raipur (India)
keshriverma@gmail.com

Ligendra Kumar Verma

Department of Computer Science
Raipur Institute of Technology
Raipur, India
ligendra@rediffmail.com

Ajay Dewangan

Vivekananda College of Technology,
Raipur
ajaydewangan1212@gmail.com

Abstract— Image classification is one of the challenging problems of image processing. This paper proposes a novel method for classification of image data. The paper emphasis on major classification methods and some extension on existing decision tree induction algorithm[4].

Keywords-Image Classification, Feature Extraction

Categories and Subject Descriptors

I.4.2 [Digital Image Processing]: Classification

I. INTRODUCTION

In various computer vision applications widely used to identify the same class of image among the set of different images. In order to classify the images it uses different features like color, texture, shape. All algorithms use the three major tasks in order to classify the images.

- Feature extraction
- Feature Selection
- Classification

Feature extraction is the process of generating properties of images to be used in the selection and classification task. Feature selection process selects the optimal number of features required to classify the images and classification phase defines the class label of any image.

Ryszard [1] has defined the different features as follows.

General features: Application independent features such as color, texture, and shape. According to the abstraction level, they can be further divided into:

- Pixel-level features: Features calculated at each pixel, e.g. color, location.
- Local features: Features calculated over the results of subdivision of the image band on image segmentation or edge detection.
- Global features: Features calculated over the entire image or just regular sub-area of an image.

a. Color feature extraction

Color texture feature extraction problem can be defined as let $\{F(x, y); x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$ be a two-dimensional image pixel array. For color images $F(x, y)$ denotes the color value at pixel (x, y) i.e., $F(x, y) = \{F_R(x, y), F_G(x, y), F_B(x, y)\}$. For black and white images, $F(x, y)$ denotes the grayscale intensity value of pixel (x, y) . In some medical science application color is one of the important feature to diagnose the disease like Tongue cancer, Skin cancer etc.

b. Shape Feature –

Second important features of images are shape. Identification of different shape is another problem of pattern recognition. The shape features is useful in satellite image processing identification of Road ways, Stadium and other important object which shapes are predefined and not dependent on dataset.

c. Texture Features

Texture feature is one of the important features for identifying objects or regions of interest in images. Haralick et. al [2] first time proposed the classification based on texture features.

In this Work intensity histogram features and Gray Level Co-Occurrence Matrix (GLCM) features are extracted.

Texture feature calculation uses the contents of GLCM to give the variation in intensity level. Texture feature calculation uses the contents of the GLCM The features that can be extracted are explained in [3]:

Homogeneity	$\sum_{i,j} \frac{p(i, j)}{1 + i - j }$
-------------	--

Entropy	$\sum_{i,j=0}^{N-1} -\ln(P_{i,j})P_{i,j}$
Contrast	$\sum_{i,j} i-j ^2 P(i,j)$
Correlation	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j}$
Energy	$\sum_{i,j} p(i,j)^2$
Sum Mean	$\frac{1}{2} \sum_i^M \sum_j^N (i * P[i,j] + P[i,j])$
Variance	$\frac{1}{2} \sum_i^M \sum_j^N ((i-\mu)^2 P[i,j] + (j-\mu)^2 P[i,j])$
Maximum Probability	$Max_{i,j}^{M,N} P[i,j]$
Inverse Difference Moment	$\sum_i^M \sum_j^N \frac{P[i,j]}{ i-j ^K}$
Cluster Tendency	$\sum_i^M \sum_j^N (i+j-2\mu)^k P[i,j]$
Correlation	$\sum_i^M \sum_j^N \frac{(i-\mu)(j-\mu)^2 P[i,j]}{\sigma^2}$

Where P is normalized co-occurrence matrix (i,j) is the pair of gray level intensities and M x N is the size of co-occurrence matrix

D. Classification

The main objective of automation is classification of objects on different class label based on their feature extracted from objects. . There are two type of classification 1. Binary classification 2. Multilabel classification

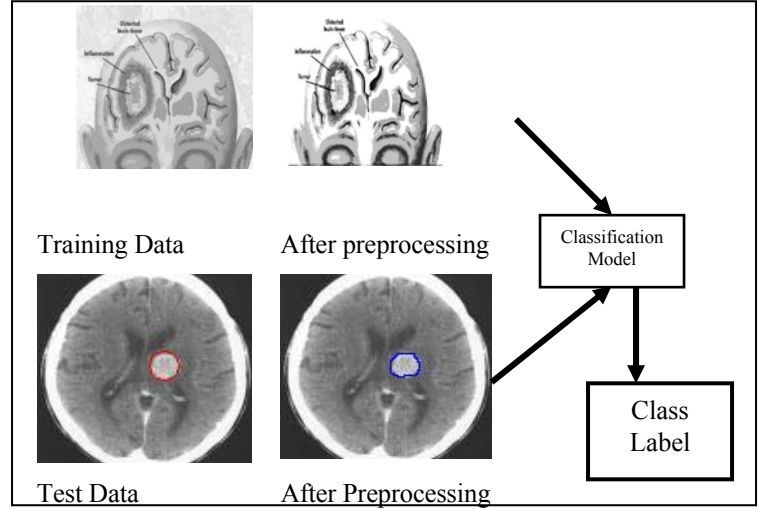
The problem of binary classification is defined as:

Input: a set of m examples (x^j, y^j), $j = 1, 2, \dots, m$ (the learning set) sampled from some distribution D, where $x^j \in R^n$ and $y^j \in \{-1, +1\}$. The i-th component of x^j, y^j is termed feature i.

Output: a function $f: R^n \rightarrow \{-1, +1\}$ which classifies “well” additional samples $\{x^k\}$ sampled from the same distribution D.

In the rest of the scribe X_i will denote the i-th feature and x^j

i the value of feature i in the jth sample. If $y^j = -1$ the sample will be referred to as a “negative sample”, and if $y^j = +1$ it will be referred to as a “positive sample”.



.Fig. 1. Classification Model

II. PROPOSED APPROACH

There are various algorithm for classify the dataset like support vector Machine, Neural Network, Genetic Algorithm, Independent Component Analysis and others for classify normal or complex data. In this paper we propose decision tree induction algorithm to classify the image data. A decision tree is a class discriminator that recursively partitions the training set until each partition consist entirely or dominantly of examples from one class. Each leaf node of the tree contains one or more attributes and determines how the data is partitioned.

A decision tree classifier is built in two phases [4] growing phase followed by pruning phase. In growth phase the tree is built by recursively partitioning the data until each partition is either “pure” or sufficiently small.

The algorithm for building tree –

Procedure buildTree (S)

- step(1). Initialize root node using dataset S
- step(2). Initialize queue Q to contain root node
- step(3). While Q is not empty do {
- step(4). dequeue the first node N in Q
- step(5). if N is not pure {
- step(6). for each attribute A
- step(7). Evaluate splits on attributes A
- step(8). Use best split to split node N into N_1 and N_2
- step(9). Append N_1 and N_2 to Q
- step(10). }
- step(11). }
- step(12).

Decision tree classifier is a mode of the data that encodes the distribution of the class label in terms of the predictor attributes, which is presented in the form of a cyclic graph a tree. The root of the tree does not have any incoming edges. Every other node has exactly one incoming edge and zero or more outgoing edge. If a node “n” has no child node than n is called leaf node. Otherwise n is called internal node. Each leaf node is labeled with one class labeled. Each internal node is labeled with one predictor attribute called the splitting attribute. Each edge “e” originating from an internal node n has a predicates “q” associated with it where “q” involves only the splitting attribute of n. The set of predicate P on the outgoing edges of an internal node must be non-overlapping and exhaustive. A set of predicate P is non-overlapping if the conjunction of any two predicates in P evaluates to false. A set of predicates P is *exhaustive* if the disjunction of all predicates P is evaluated to true. We will call the set of predicates on the outgoing edges of an internal node the *splitting predicates of n*.

Some extracted features of images are defined in table [1]. Table 1. shows the different features of 10 different Medical images(MRI and CT Scan). Each of them have a distinct values. After categorical conversion of data after calculation of median value.

S.No.	Features						Class label
	Mean	Variance	RMS	Mode	Median	entropy	
1	83.00489	77.55259	113.59672	0.00000	68.00000	7.072181	N
2	192.49353	40.60893	196.73038	212.0000	208.0000	4.52083	Y
3	76.48813	77.400375	108.81751	254.00000	111.0000	5.755.1	Y
4	54.29624	66.71162	86.01467	0.000000	120.0000	5.3269	N
5	84.5975	67.7291	108.32281	10.00000	65.00000	7.3603	N
6	65.35561	67.4.259	93.88538	0.000000	99.00000	6.12887	N
7	94.89065	66.6749	195.46611	9.000000	111.0000	7.1913	Y
8	138.4066	59.24549	150.55376	74.00000	133.0000	6.83289	Y
9	74.94568	68.76751	1.1.67343	9.000000	68.00000	7.14518	Y
10	65.8125	69.27354	95.5516	3.000000	43.00000	6.33066	Y

After preprocessing of above dataset the table [1] is transformed in table[2]. It converts the numerical data into categorical dataset.

Table 2. Conversion of Numerical Data into Categorical Dataset

S.N o.	Features						Class label
	Mean	Varian ce	RMS	Mode	Media n	entro py	
1	LOW	HIGH	LOW	LOW	LOW	HIGH	N
2	HIGH	LOW	HIGH	HIGH	HIGH	LOW	Y
3	LOW	HIGH	LOW	HIGH	LOW	HIGH	Y

4	LOW	HIGH	LOW	LOW	LOW	HIGH	N
5	LOW	HIGH	LOW	LOW	LOW	LOW	N
6	LOW	HIGH	LOW	LOW	LOW	HIGH	N
7	LOW	HIGH	HIGH	LOW	LOW	HIGH	Y
8	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	Y
9	LOW	HIGH	LOW	LOW	LOW	HIGH	Y
10	LOW	HIGH	LOW	LOW	LOW	HIGH	Y

The above table can be converted into decision tree as follows.

Construction of decision tree [4] as quinlan algorithm works as follows.

A. Entropy Formulae

Entropy, a measure from information theory, characterizes the (im) purity, or homogeneity, of an arbitrary collection of examples.

Given:

- n_b , the number of instances in branch b .
- n_{bc} , the number of instances in branch b of class c .
Of course, n_{bc} is less than or equal to n_b
- n_i , the total number of instances in all branches.

Probability

P_b = Probability in instance on a branch b is positive

$$= \frac{\text{Number of positive instance on branch}}{\text{total number of instance on branch}}$$

$$= n_{bc} / n_b$$

- If all the instances on the branch are positive, then $P_b = 1$ (homogeneous positive)
- If all the instances on the branch are negative, then $P_b = 0$ (homogeneous negative)

B. Entropy

$$\text{Entropy} = \sum_c - \left(\frac{n_{bc}}{n_b} \right) \log_2 \left(\frac{n_{bc}}{n_b} \right)$$

As we move from perfect balance and perfect Homogeneity, entropy varies smoothly between zero and one.

- The entropy is zero when the set is perfectly homogeneous.

- The entropy is one when the set is perfectly inhomogeneous.

Average Entropy

$$\text{Average Entropy} = \sum_b \left(\frac{n_b}{n_t} \right) \times \left[\sum_c - \left(\frac{n_{bc}}{n_b} \right) \log_2 \left(\frac{n_{bc}}{n_b} \right) \right]$$

C. Information Gain

The expected reduction in entropy caused by partitioning the data according to an attributes is defined as [4]

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Value (A) is the set of all possible values for attribute A, and S_v is the subset of S for which attribute A has value v.

After applying above we get the value for information gain as follows for above dataset.

Table 3 : Information Gain for different attributes.

Attribute	I(G)
Mean	2.844
Variance	2.41
RMS	2.43
Mode	2.8482
Median	2.80
Entropy	2.84

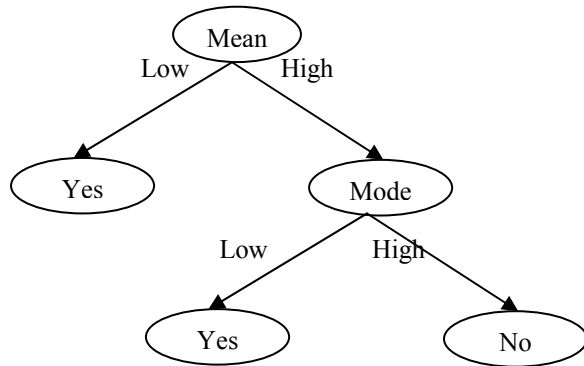


Fig. 1 . Decision tree for Brain Tumor dataset.

Conversion of decision tree into rules –

Rule 1 : If Mean = “High” Class label = “yes”

Rule 2 : IF Mean = “Low” and Mode = “High” class label = “Yes”

Rule 3: IF Mean = “Low” and Mode = “Low” class label = “NO”

III CONCLUSION AND FUTURE WORK

In this paper we described the different texture features of images. We proposed a novel method decision tree induction based on quinlan [3] theory to classify digital images. The experiment is performed on matlab 2010a in order to extract the features of brain cancer data. We have taken 10 images and 6 attributes for feature extraction. The attributes for feature extraction are mean, variance, RMS, mode, median and entropy. We drawn the decision tree based on highest information gain of data. Furthermore the decision tree is converted into rules. The future work is to obtain all the 14 features of haralick[2] theorem and the experimentation on large dataset, so that we can obtain some strong rule for disease diagnosis.

REFERENCES

- [1]. Biometrics Systems. INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING. Issue 1 Volume 1 page 6-16,2007.
- [2]. Haralick, R.M. Shanmugan K. and Dinstein I, 1973. Texture Features for image classification IEEE Transaction on System amn and Cybermatics, Vol. 3 NO. 6 pp 610-621.
- [3]. Quinlan J. Ross Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufman, 1993. [Qui93] J. Ross Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufman, 1993.
- [4]. Quinlan J. R. and R. L. Rivest. Inferring decision trees using minimum description length principle. Information and Computation, 1989.
- [5]. Feature Selection: Evaluation, Application, and Small Sample Performance. Jain, Anil and Zongker, Douglas. 1997, IEEE Trans. Pattern Anal. Machine Intell, Vol. 19, pp. 153-158.
- [6]. P. Geurts, Contributions to decision tree induction: bias/variance tradeoff and time series classification, Phd. thesis, Department of Electrical Engineering and Computer Science, University of Li'ège, May 2002.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Mrs Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Mr. P. Vasant, University Technology Petronas, Malaysia
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Mr. Praveen Ranjan Srivastava, BITS PILANI, India
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia

Mr. Tirthankar Gayen, IIT Kharagpur, India
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Mr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Mr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan
Prof. Syed S. Rizvi, University of Bridgeport, USA
Mr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Mr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Mr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Mr. S. Mehta, Inha University, Korea
Mr. Dilip Kumar S.M, University Visvesvaraya College of Engineering (UVCE), Bangalore University,
Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Mr. Saqib Saeed, University of Siegen, Germany
Mr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Mrs.J.Komala Lakshmi, SNR Sons College, Computer Science, India
Mr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, Faculty of Computer Science and Information Technology, University of
Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Mr. M. Azath, Anna University, India
Mr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Mr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr Suresh Jain, Professor (on leave), Institute of Engineering & Technology, Devi Ahilya University, Indore
(MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Mr. Hanumanthappa. J. University of Mysore, India
Mr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Mr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria

Mr. Santosh K. Pandey, Department of Information Technology, The Institute of Chartered Accountants of India

Dr. P. Vasant, Power Control Optimization, Malaysia

Dr. Petr Ivankov, Automatika - S, Russian Federation

Dr. Utkarsh Seetha, Data Infosys Limited, India

Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal

Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore

Assist. Prof. A. Neela madheswari, Anna university, India

Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India

Mr. Kamanashis Biswas, Daffodil International University, Bangladesh

Dr. Atul Gonsai, Saurashtra University, Gujarat, India

Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand

Mrs. G. Nalini Priya, Anna University, Chennai

Dr. P. Subashini, Avinashilingam University for Women, India

Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat

Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal

Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India

Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India

Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah

Mr. Nitin Bhatia, DAV College, India

Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India

Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia

Assist. Prof. Sonal Chawla, Panjab University, India

Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India

Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia

Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia

Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India

Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France

Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India

Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology, Durban,South Africa

Prof. Mydhili K Nair, M S Ramaiah Institute of Technology(M.S.R.I.T), Affiliated to Visweswaraiah Technological University, Bangalore, India

M. Prabu, Adhiyamaan College of Engineering/Anna University, India

Mr. Swakkhar Shatabda, Department of Computer Science and Engineering, United International University, Bangladesh

Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan

Mr. H. Abdul Shabeer, I-Nautix Technologies,Chennai, India

Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India

Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, : Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, College of Computer Science and Engineering, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institutiion of Engg. & Tech. CHD, India
Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India
Dr. C. Arun, Anna University, India

Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology
Subhabrata Barman, Haldia Institute of Technology, West Bengal
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand
Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
Mr. Serguei A. Mokhov, Concordia University, Canada
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA
Dr. S. Karthik, SNS College of Technology, India
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain
Mr. A.D.Potgantwar, Pune University, India
Dr. Himanshu Aggarwal, Punjabi University, India
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India
Dr. K.L. Shunmuganathan, R.M.K Engg College, Kavaraipeitai, Chennai
Dr. Prasant Kumar Pattnaik, KIST, India.
Dr. Ch. Aswani Kumar, VIT University, India
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan
Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA
Mr. R. Jagadeesh Kannan, RMK Engineering College, India
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh

Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mohammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET , Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded , India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhanian University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow ,UP India
Dr. V V S S S Balaram, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy. P, KITS , Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A.Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen , Aberystwyth University, UK
Dr . Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praaksh Singh , Academy of Technology, India
Dr. Ritu Soni, GNG College, India
Dr . Mahendra Kumar , Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, Lakshmi Narayan College of Tech.(LNCT)Bhopal India

Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath , ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan
Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwaocha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto , Singapore
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhania University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhania University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India

Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India
Dr. Hanan Elazhary, Electronics Research Institute, Egypt
Dr. Hosam I. Faiq, USM, Malaysia
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India
Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India
Prof Anupam Choudhary, Bhilai School Of Engg., Bhilai (C.G.), India
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.
Dr. Kasarapu Ramani, JNT University, Anantapur, India
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India
Dr. C G Ravichandran, R V S College of Engineering and Technology, India
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India
Dr. Nikolai Stoianov, Defense Institute, Bulgaria
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh
Mr. Hemanta Kumar Kalita, TATA Consultancy Services (TCS), India
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia
Mr. B. Muthu Kumar, Kathir College Of Engineering, Coimbatore
Dr. Nighat Mir, Effat University, Saudi Arabia
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India
Mr. P. Sivakumar, Anna university, Chennai, India

Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia
Mr. Nikhil Patrick Lobo, CADES, India
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India
Assist. Prof. Vishal Bharti, DCE, Gurgaon
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India
Mr. Hamed Taherdoost, Tehran, Iran
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran
Mr. Shantanu Pal, University of Calcutta, India
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria
Mr. P. Mahalingam, Caledonian College of Engineering, Oman
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt
Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India
Mr. Muhammad Asad, University of Engineering and Technology Taxila, Pakistan
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran
Prof. S. V. Nagaraj, RMK Engineering College, India
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India
Mr. Sunil Taneja, Kurukshetra University, India
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia
Dr. Yaduvir Singh, Thapar University, India
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia

Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor Zarei, UAE Inventors Association, UAE
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India
Prof. Anant J Umbarkar, Walchand College of Engg., India
Assist. Prof. B. Bharathi, Sathyabama University, India
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore
Prof. Walid Moudani, Lebanese University, Lebanon
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India
Associate Prof. Dr. Manuj Darbari, BBD University, India
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India
Dr. Abhay Bansal, Amity School of Engineering & Technology, India
Ms. Sumita Mishra, Amity School of Engineering and Technology, India
Professor S. Viswanadha Raju, JNT University Hyderabad, India
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia
Mr. Adri Jovin, Anna University of Technology, Coimbatore, India
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia
Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India
Mr. Shervan Fekri Ershad, Shiraz International University, Iran
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India
Ms. Sarla More, UIT, RGTU, Bhopal, India
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India
Dr. M. N. Giri Prasad, JNTUCE,Pulivendula, A.P., India

CALL FOR PAPERS
International Journal of Computer Science and Information Security
IJCSIS 2011
ISSN: 1947-5500
<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2011
ISSN 1947 5500